# Instrumental variable estimation in the presence of many moment conditions

AUTHOR(S):

Okui, Ryo

CITATION:

Okui, Ryo. Instrumental variable estimation in the presence of many moment conditions. Journal of Econometrics 2011, 165(1): 70-86

ISSUE DATE:

2011-11

URL:

http://hdl.handle.net/2433/147964

# Instrumental variable estimation in the presence of many moment conditions*

Ryo Okui†
Kyoto University

November 9, 2010

**Abstract**

This paper develops shrinkage methods for addressing the "many instruments" problem in the context of instrumental variable estimation. It has been observed that instrumental variable estimators may behave poorly if the number of instruments is large. This problem can be addressed by shrinking the influence of a subset of instrumental variables. The procedure can be understood as a two-step process of shrinking some of the OLS coefficient estimates from the regression of the endogenous variables on the instruments, then using the predicted values of the endogenous variables (based on the shrunk coefficient estimates) as the instruments. The shrinkage parameter is chosen to minimize the asymptotic mean square error. The optimal shrinkage parameter has a closed form, which makes it easy to implement. A Monte Carlo study shows that the shrinkage method works well and performs better in many situations than do existing instrument selection procedures.

**Keywords**: TSLS, LIML, Shrinkage estimator, Instrumental variables.
**JEL classification**: C21, C31

*A previous version of this paper was circulated under the title "Shrinkage methods for instrumental variable estimation." This study is part of the author's dissertation project at the University of Pennsylvania. The author gratefully acknowledges the hospitality of Yale University, where part of this paper was written. The author thanks his advisor, Yuichi Kitamura, for his helpful guidance. Gregory Kordas, Petra Todd and Frank Schorfheide also provided helpful supervision. The author obtained valuable comments from two anonymous referees, Dylan Small, Yoshihiko Nishiyama, Sokbae Lee, Donald Andrews, Peter Phillips, Taisuke Otsu, Whitney Newey, Naoto Kunitomo, members of the Penn Empirical Micro Discussion Group, and seminar participants at various institutes. The author is also indebted to Claire Lim, Shalini Roy and Jason Harburger. The author acknowledges financial support from the Research Grants Council of Hong Kong under Project No. HKUST643907. All remaining errors are the author's.

†Institute of Economic Research, Kyoto University, Yoshida-Honmachi, Sakyo, Kyoto, Kyoto, 606-8501, Japan, tel:+81-75-753-7191, e-mail: okui@kier.kyoto-u.ac.jp

# 1 Introduction

This paper proposes new solutions to the problem of instrumental variable (IV, hereafter) estimation in the presence of many instruments. In this situation, we can estimate the model and make some inferences using a minimal subset of instruments. However, with a small number of instruments, we lose efficiency, which results in relatively large standard errors. We might try to increase the number of instruments in order to reduce the standard error of the estimate. It turns out that this approach may be misleading in finite samples. An IV estimator with many instruments may behave poorly and can be sensitive to the number of instruments. In particular, the two-stage least squares (TSLS, hereafter) estimator generates a bias the order of which is proportional to the number of instruments (e.g., see Kunitomo (1980), Morimune (1983), or Bekker (1994)).[1] An example where this problem occurs in empirical work is the paper by Angrist and Krueger (1991).[2] Bound, Jaeger and Baker (1996) illustrate how the problem of many instruments arises in Angrist and Krueger's (1991) work.[3]

Existing solutions to the "many instruments" problem usually involve instrument selection. Donald and Newey (2001) propose minimizing the asymptotic

[1]Morimune (1985) is a good reference that summarizes the development of the researches on this topic until the middle of 1980's. Unfortunately, the book is available only in Japanese.

[2]They estimate the return to an additional year of schooling. They show that quarter-of-birth can be an instrument to years of schooling. Their set of instruments includes quarter-birth variables and their interactions with year-of-birth and state-of-birth variables. The number of (excluded) instruments is 180 in one of their specifications.

[3]Even though Bound, Jaeger and Baker (1996) emphasize the weak instrument problem, Table 1 in their paper indicates that Angrist and Krueger's (1991) data do not suffer from the bias of the TSLS estimator if we use a minimal subset of instruments. See also Hansen, Hausman and Newey (2008). Actually, there are two problems: one is the "many instruments" problem and the other is that the additional instruments are weak. This paper focuses on the "many instruments" problem. Chao and Swanson (2005) and Stock and Yogo (2005) discuss the consequences of a large number of weak instruments, and Anderson, Kunitomo and Matsushita (2008) provide extensive simulation results.

mean square error as a criterion for choosing the number of instruments. Small (2002) proposes a criterion function motivated by the Akaike Information Criteria for choosing instruments. Hall and Peixe (2003) also consider another information criterion for instrument selection. Their information criterion consists of two terms. The first term is based on the canonical correlations which measure the relevance of moment conditions. The second term penalizes the number of moment conditions.

This paper introduces a new procedure for IV estimation based on shrinkage methods. That is, we reconstruct the estimating equation of an IV estimator, which is a weighted sum of sample moment conditions, by shrinking some elements of the weighting vector. This idea can also be interpreted as shrinking part of the OLS coefficient estimates from the regression of the endogenous variables on the instruments and then using the predicted values of the endogenous variables, based on the shrunk coefficient estimates, as the instruments.

One nontrivial question is how to choose the shrinkage parameter. We propose to choose the shrinkage parameter by minimizing the Nagar (1959)-type approximation of the mean square error. The optimal shrinkage parameter has a closed form, which makes it easy to implement. Alternatively, we may consider choosing the shrinkage parameter in a similar way as the well-known James–Stein estimator. However, the James–Stein shrinkage rule is not optimal, and in shrinkage TSLS estimation, there is a crucial difference between these two. Note that the James–Stein shrinkage rule has just an order-$K$ term where $K$ is the number of instruments, however; the optimal shrinkage parameter has an order-$K^2$ term. The shrinkage parameter given by the James–Stein shrinkage rule is larger than desired when the number of instruments is large. This shows the importance of the mean square error calculation in choosing the shrinkage parameter.

In the statistical literature, it has been observed that shrinkage methods perform well, and, moreover, they often work better than selection methods (e.g., see Hastie, Tibshirani and Friedman (2001), Section 3.4.5). The key decision involved

in selection methods is to select which instruments to discard. Even though we alleviate the many-instruments problem by doing so, we also ignore the information that the discarded instruments might reveal. On the other hand, shrinkage methods not only mitigate the many-instruments problem but also enable the use of the information that is lost by discarding variables. Shrinkage procedures can become excellent alternatives to selection methods in IV estimation.

A limitation of the shrinkage method proposed here is that it requires us to specify the set of "main" IVs which are *a priori* known to be strong. While this requirement may be restrictive in some applications, there are situations in which it is possible to specify the set of "main" IVs in a natural way. For example, Angrist and Krueger (1991) use quarter–of–birth variables and their interactions with year–of–birth or state–of–birth variables as instruments. In this case, the quarter–of–birth variables may be considered as "main" instruments and the interactions may be considered as other instruments. We note that selection methods such as those of Donald and Newey (2001) typically require a different assumption that an ordering of instruments is prespecified to make them computationally feasible and to justify the method theoretically.

Even though there is hardly any literature that explicitly considers the application of shrinkage methods in IV estimations, Chamberlain and Imbens (2004) consider a procedure, called random effect quasi-maximum likelihood (REQML), which could be categorized as a shrinkage method. They impose a random effect structure on the coefficients in the regression of the endogenous variable on instruments and then maximize the likelihood that takes the random effect structure into account. REQML has several attractive features, such as being interpretable as a Bayes procedure. However, extending the idea of their paper to different settings may not be trivial. For example, the appropriate way to construct the likelihood function of a conditional moment restriction model with conditional heteroskedasticity is not necessarily clear. Moreover, it is also not clear what the appropriate way would be to impose a random effect structure in such a model. The procedure

presented here can be extended to different models, as shown by Okui (2005), to consider conditional moment restriction models and dynamic panel data models. Another limitation of REQML is that handling a situation with multiple endogenous regressors is not straight-forward. On the other hand, the method considered in this paper is applicable to such a situation. Finally, we derive an approximation of the mean square error of the estimator and choose the shrinkage parameter by minimizing the approximate mean square error, while Chamberlain and Imbens (2004) do not consider the mean square error of the estimator. The kernel-weighted GMM in ARMA models by Kuersteiner (2002) is also related to the ideas explored here.[4] Another related paper is Carrasco (2008). Her idea is different from the one considered here. Her approach involves regularization of the inverse of the covariance matrix of instruments while our approach is to shrink some of the coefficient estimates in the first stage regression.

The rest of this paper is organized as follows. The next section introduces the shrinkage TSLS estimator, explains the motivation for the method and presents the theoretical results. Section 3 proposes the shrinkage limited information maximum likelihood estimator. Results from Monte Carlo experiments are included in Section 4. Discussions and possible extensions are presented in Section 5.

We use the following notation throughout the paper. For a sequence of vectors, $\{A_i\}$, we define $A$ as $A = (A_1', A_2', \ldots, A_n')'$. For a matrix $A$, we define $||A|| =$

[4]We note that there is a pair of kernel functions and bandwidths under which the kernel-weighted GMM and the shrinkage TSLS are equivalent. They are $K(u) = 1$, for $|u| < c$ and $K(u) = s$ for $|u| \geq c$, where $s$ is the shrinkage parameter and $c$ is equal to the ratio of the number of main instruments and the total number of instruments, and the bandwidth is equal to the total number of instruments. We note that the choices of bandwidth and shrinkage parameter are not equivalent. Roughly speaking, the shrinkage TSLS chooses the kernel function given the bandwidth, whereas the kernel-weighted GMM chooses the bandwidth given the kernel function. Thus, there is a fundamental difference between the kernel-weighted GMM and shrinkage methods. The kernel-weighted GMM can be regarded as a way to exploit all information from the order of instruments that is clear in ARMA models, while this paper implicitly considers situations where the order is not clear.

$\sqrt{tr(A'A)}$ (the usual Euclidean norm), and $P_A = A(A'A)^{-1}A'$.

# 2 The Shrinkage TSLS Estimator

## 2.1 Model and Procedure

Following Donald and Newey (2001), we consider the model:

$$\begin{aligned} y_i &= Y_i'\gamma + x_{1i}'\beta + \epsilon_i = W_i'\delta + \epsilon_i \\ W_i &= \begin{pmatrix} Y_i \\ x_{1i} \end{pmatrix} = f(x_i) + u_i = \begin{pmatrix} E(Y_i|x_i) \\ x_{1i} \end{pmatrix} + \begin{pmatrix} \eta_i \\ 0 \end{pmatrix}, \quad i = 1, \ldots, N, \end{aligned}$$

where $y_i$ is a scalar outcome variable, $Y_i$ is a $d_1 \times 1$ vector of endogenous variables, $x_i$ is a vector of exogenous variables, $\epsilon_i$ and $u_i$ are unobserved random variables with second moments, which do not depend on $x_i$, and $f$ is an unknown function of $x_i$. Let $f_i = f(x_i)$. The set of instruments is $(X_i', \bar{Z}_i')$. $X_i$ is an $m \times 1$ vector of main instruments and $\bar{Z}_i$ is a $K \times 1$ vector of other instruments. They are functions of $x_i$. The included exogenous variable, $x_{1i}$, is a part of $X_i$. We employ this semiparametric structure because it allows us to analyze easily the model with many instruments. Another reason is that this paper intends to compare instrument selection methods and shrinkage methods, and, to this end, it would be better to have the same structure as used in Donald and Newey (2001) to present a selection method that will be compared with shrinkage procedures in the Monte Carlo section.

In the current model, the asymptotic variance of a $\sqrt{N}$-consistent regular estimator cannot be smaller than $\sigma^2\bar{H}^{-1}$, where $\sigma^2 = E(\epsilon_i^2|x_i)$ and $\bar{H} = E(f_i f_i')$ (Chamberlain (1987)). It can be achieved if $f_i$ can be written as a linear combination of the instruments. Likewise, if there is a linear combination of the instruments that is close to $f_i$, then the asymptotic variance of the IV estimator is small. This observation implies that using many instruments is desirable in terms of asymptotic variance. However, an IV estimator with many instruments may behave poorly in

finite samples and can be sensitive to the number of instruments. Furthermore, if a set of instruments can approximate $f_i$ well, then adding more instruments is not helpful to reducing the asymptotic variance since it cannot be smaller than $\sigma^2 \bar{H}^{-1}$. It is, therefore, important to consider how to handle a large number of instruments.

We consider a situation similar to that of Chamberlain and Imbens (2004) where we have two sets of instruments, $X$ and $\bar{Z}$. Among the IVs, we typically have "main" instruments, which guarantee the identification of the parameter, $\delta$, and are more important for estimation than the other instruments. We denote these "main" instruments as $X$. We consider shrinking the effect of $\bar{Z}$ on the estimation of $\delta$. The meaning of "main" can differ among situations. For example, suppose that we consider a (possibly misspecified) linear (in parameters) model for the relationship between the endogenous regressors and instruments, as in West, Wong and Anatolyev (2009). The main instruments (i.e., $X$) in this case would be the terms appearing in the model we specify and the other instruments (i.e., $\bar{Z}$) are other functions of the instruments. Another example could be the case where a number of instruments are generated by multiplying the main instruments by regional dummies or time dummies. For instance, as discussed in Section 1, the quarter of birth variables may be considered as "main" instruments in the case of Angrist and Krueger (1991).

Note that we are able to estimate $\delta$ using only those main instruments if the number of the main instruments is larger than the number of the endogenous variables. However, such an estimate may have a large standard error. Even though using more instruments is a way to reduce the standard error of the estimate, it is commonly observed that IV estimators with many instruments behave poorly (e.g., Morimune (1983) and Bound, Jaeger and Baker (1996)). The shrinkage TSLS estimator is introduced to address this "many instruments" problem. In this section, the shrinkage TSLS estimator is discussed. The shrinkage LIML estimator is discussed in the next section.

Now, we describe the procedure. Let $Z = (I - P_X)\bar{Z}$ so that $X'Z = 0$. It is important to note that $Z$ in our discussion may not be the matrix of the instruments itself but the orthogonalized one in applications. The TSLS estimator of $\delta$ is the solution to

$$W'P_X(y - W\delta) + W'P_{(X,Z)}(y - W\delta) = 0.$$

The shrinkage TSLS estimator, $\hat{\delta}_{tsls,s}$, is defined as the solution to

$$\begin{aligned} & (1-s)W'P_X(y - W\delta) + sW'P_{(X,Z)}(y - W\delta) \\ = \; & W'P_X(y - W\delta) + sW'P_Z(y - W\delta) = 0, \end{aligned}$$

and it is:

$$\hat{\delta}_{tsls,s} = (W'P^sW)^{-1}W'P^sy,$$

for a shrinkage parameter, $s$, where $P^s = (1-s)P_X + sP_{(X,Z)} = P_X + sP_Z$. The shrinkage TSLS estimator is obtained by solving a weighted average of the estimating equation for the TSLS using only the main instruments and that using all instruments. By introducing the shrinkage parameter, $s$, we can reduce the effect of adding $Z$ into the set of instruments. The shrinkage parameter, $s$, lies between 0 and 1; the choice $s = 0$ leads to the TSLS estimator using only $X$ and likewise setting $s = 1$ yields the TSLS estimator using all the instruments. A more detailed discussion is found in the next subsection.

To operationalize this procedure, a method for choosing $s$ is needed. We recommend the following choice of $s$ because it is an estimator of the shrinkage parameter that minimizes the asymptotic mean square error (see Section 2.3):

$$\hat{s}^* = \frac{\hat{\sigma}_\epsilon^2 \dfrac{\hat{\lambda}'\hat{H}^{-1}W'P_ZW\hat{H}^{-1}\hat{\lambda}}{N}}{\hat{\lambda}'\hat{H}^{-1}\hat{\sigma}_{u\epsilon}\hat{\sigma}_{u\epsilon}'\hat{H}^{-1}\hat{\lambda}\dfrac{K^2}{N} + \hat{\sigma}_\epsilon^2 \dfrac{\hat{\lambda}'\hat{H}^{-1}W'P_ZW\hat{H}^{-1}\hat{\lambda}}{N}}, \tag{1}$$

where $\hat{\lambda}$ is the (possibly estimated) weighing vector chosen by the researcher, $\hat{\sigma}_\epsilon^2$ and $\hat{\sigma}_{u\epsilon}$ are the estimates of $\sigma_\epsilon^2 = E(\epsilon_i^2)$ and $\sigma_{u\epsilon} = E(u_i\epsilon_i)$ based on the residuals

from a preliminary estimation and $\hat{H} = W'(P_X + P_Z)W/N$, which is an estimate of $H = f'f/N$ (i.e., the first-order asymptotic variance). Note that the number of instrumental variables should increase with the sample size in order to estimate $H$.[5]

## 2.2 Theoretical Results

We demonstrate the asymptotic properties of the shrinkage TSLS under the following assumptions, which are similar to those imposed in Donald and Newey (2001):

**Assumption 1.** *$\{y_i, W_i, x_i\}$ are i.i.d., $E(\epsilon_i^2|x_i) = \sigma_\epsilon^2 > 0$, and $E(||\eta_i||^4|x_i)$ and $E(|\epsilon_i|^4|x_i)$ are bounded.*

**Assumption 2.** *(i) $\bar{H} \equiv E(f_i f_i')$ exists and is nonsingular. (ii) there exists $\pi_K$ such that $E(||f(x) - \pi_K(X', \bar{Z}')||^2) \to 0$ as $K \to \infty$.*

**Assumption 3.** *(i) $E\{(\epsilon, u')'(\epsilon, u')|x_i\}$ is constant. (ii) $(X, Z)'(X, Z)$ is nonsingular with probability one. (iii) $\max_{i \le N} P_{Z,ii} \to_p 0$. (iv) $f_i$ is bounded.*

Assumption 1 imposes restrictions on the moments of the random variables in the model, which are standard in the literature. Assumption 2(i) guarantees the identification of the parameter $\delta$. With Assumption 2(ii), we have the asymptotic variance of the TSLS estimator or the shrinkage TSLS estimator under $K \to \infty$ is $\sigma^2 \bar{H}^{-1}$. Assumption 3(i) imposes homoskedasticity of the error terms. Assumption 3(ii) and (iii) impose restrictions on the probabilistic nature of the instruments and the rate of $K$. Assumption 3(iv) is employed for simplicity and can be relaxed at the cost of making the results and the proofs much more complicated.

The first theorem is on the consistency and the asymptotic normality of the shrinkage TSLS estimator.

[5] Note that the function $f(\cdot)$ is unknown and it cannot be written as a linear combination of a finite number of instruments in general. This is the reason that the number of instruments should increase with $N$ in order to estimate $H = f'f/N$.

**Theorem 1.** *Suppose Assumptions 1-3 are satisfied. If* $(sK)^2/N \to_p 0$ *and either* $s \to_p 1$ *or* $E(f_i Z_i') = 0$*, then* $\hat{\delta}_{tsls,s} - \delta \to_p 0$ *and* $\sqrt{N}(\hat{\delta}_{tsls,s} - \delta) \to_d N(0, \sigma_\epsilon^2 \bar{H}^{-1})$.

The condition $E(f_i Z_i) = 0$ means that $Z$ is a matrix of totally irrelevant instruments and in that case the shrinkage parameter does not need to go to 1. However, when $Z$ is relevant, the shrinkage parameter must converge to 1 in order to achieve the semiparametric efficiency bound. This theorem justifies the use of the shrinkage TSLS estimator. Unfortunately, this result also indicates that the conventional first-order asymptotic analysis is neither strong enough to investigate the effect of shrinkage, nor able to provide any guidance in choosing the shrinkage parameter, $s$. This is similar to the case of selecting the number of instruments. The first-order asymptotic results do not tell us how many instruments should be used; for this, we have to look at a higher-order expansion. Given this observation, we propose to choose the shrinkage parameter to minimize a higher-order asymptotic mean square error. The notion of the asymptotic mean square error employed here is similar to the Nagar-type asymptotic expansion (Nagar (1959)). Following Donald and Newey (2001), we approximate the mean square error, $E\{(\hat{\delta}_{tsls,s} - \delta_0)(\hat{\delta}_{tsls,s} - \delta_0)'|x\}$, by $\sigma_\epsilon^2 H^{-1} + S(s)$ where

$$N(\hat{\delta}_{tsls,s} - \delta_0)(\hat{\delta}_{tsls,s} - \delta_0)' = \hat{Q}(s) + \hat{r}(s), \quad E\{\hat{Q}(s)|x\} = \sigma_\epsilon^2 H^{-1} + S(s) + T(s),$$

$H = f'f/N$ and $\{\hat{r}(s)+T(s)\}/tr\{S(s)\} = o_p(1)$ as $K \to \infty$ and $N \to \infty$. First, we divide the $N(\hat{\delta}_{tsls,s}-\delta_0)(\hat{\delta}_{tsls,s}-\delta_0)'$ into two parts, $\hat{Q}(s)$ and $\hat{r}(s)$, and discard $\hat{r}(s)$, which goes to zero more quickly than $S(s)$ does. Then, we take the expectation of $\hat{Q}(s)$ conditional on the exogenous variable, $x$, and ignore the term $T(s)$, which goes to zero more quickly than $S(s)$ does. The term $\sigma_\epsilon^2 H^{-1}$ corresponds to the first-order asymptotic variance. Hence, $S(s)$ is the nontrivial and dominant term in the mean squared error and our goal is to find $S(s)$.

This Nagar-type approximation is popular in the IV estimation literature but not common in the shrinkage literature, which mainly focuses on exact finite sample properties. We have several reasons to investigate the Nagar approximation even

though the usual shrinkage literature does not use it. First, a finite sample parametric approach may not be very convincing because it relies on a distributional assumption. Second, an exact finite sample approach usually gives us results that are too complicated to be meaningful. The application of the Nagar approximation provides a clear result, which leads to an easily implementable procedure for choosing the optimal shrinkage parameter. Lastly, this approach makes comparison with Donald and Newey (2001) easier as they also use the Nagar expansion.

The next theorem shows the form of the mean square error under $K \to \infty$, $N \to \infty$ and an exogenous shrinkage parameter.

**Theorem 2.** *Suppose that Assumptions 1-3 are satisfied. Under* $(sK)^2/N \to 0$ *and either* $(1-s) = O_p(K^2/N)$ *or* $E(f_i Z_i') = 0$*, Nagar's decomposition holds with*

$$S(s) = H^{-1}\left\{\sigma_{u\epsilon}\sigma_{u\epsilon}'\frac{(sK)^2}{N} + \sigma_\epsilon^2\frac{f'(I-P^s)(I-P^s)f}{N}\right\}H^{-1}.$$

The appendix contains the proof, which is similar to the proof to Proposition 1 in Donald and Newey (2001). The first term in the brackets on the right-hand side of the equation corresponds to the square of the bias term. Introducing the shrinkage parameter mitigates the bias caused by using many instruments. The second term in the brackets corresponds to the second-order variance term. Note that the formula in Donald and Newey (2001) is given by setting $s = 1$, as $s = 1$ corresponds to the standard TSLS estimator.

Given this formula, our task is to find an $s$ that minimizes the mean square error of a linear combination of the estimator, $\lambda' S(s)\lambda$ ($\lambda$ may be estimated). The optimal shrinkage parameter is:

$$s^* = \frac{\sigma_\epsilon^2\dfrac{\lambda' H^{-1}f'P_Z f H^{-1}\lambda}{N}}{\lambda' H^{-1}\sigma_{u\epsilon}\sigma_{u\epsilon}'H^{-1}\lambda\dfrac{K^2}{N} + \sigma_\epsilon^2\dfrac{\lambda' H^{-1}f'P_Z f H^{-1}\lambda}{N}}.$$

This form is very intuitive: the optimal shrinkage parameter is an increasing function of a measure of the strength of the instruments, $f'P_Z f/N$, and a decreasing function of the number of instruments, $K$. The optimal shrinkage parameter lies

between 0 and 1, which is a natural parameter space for the shrinkage parameter. The event $s^* = 1$ occurs when $\sigma_{u\epsilon} = 0$. In this case, the OLS estimator is consistent, and we should make the estimator close to the OLS estimator by using all the instruments.

The standard case is $f'P_Z f/N \to_p c > 0$ and $s^* \to_p 1$. This means that if $Z$ is a valid instrument, then asymptotically we do not shrink and achieve semiparametric efficiency. If $f'P_Z f/K^2 \to_p 0$, which occurs when $Z$ is an irrelevant instrument, $s^* \to_p 0$. We can defend against completely weak instruments by introducing the shrinkage parameter. The weak instruments case in the Staiger and Stock (1997) (see also Chao and Swanson (2005)) sense occurs when $f'P_Z f/K^2 \to_p c > 0$. Then, $s^* \to_p \bar{s}$ where $0 < \bar{s} < 1$. Even though we do not consider this case formally, we conjecture that the shrinkage TSLS can even utilize information from weak instruments.

The optimal shrinkage parameter has a $K^2$-order term. This is the main difference of this shrinkage rule compared with that of James–Stein, which just has a $K$-order term.[6] This might imply that if we employ the James–Stein shrinkage rule naively, we shrink the effect of the instruments less than desired when the number of instruments is large. This observation indicates the importance of choosing the shrinkage parameter based on the asymptotic mean square error.

If there is only one endogenous variable, or, in other words, $Y_i$ is a scalar, the choice of $\lambda$ does not affect the optimal shrinkage parameter, which is

$$s^* = \frac{\sigma_\epsilon^2 \dfrac{\bar{Y}' P_Z \bar{Y}}{N}}{\sigma_{\eta\epsilon}^2 \dfrac{K^2}{N} + \sigma_\epsilon^2 \dfrac{\bar{Y}' P_Z \bar{Y}}{N}},$$

where $\bar{Y} = (E(Y_1|x_1), \ldots, E(Y_N|x_N))'$.

The optimal shrinkage parameter depends on the unknown parameters. A natural estimator of the optimal shrinkage parameter is given by (1), and the

[6] Suppose that there is only one endogenous regressor. Then, the James-Stein shrinkage rule for the first-stage regression gives $\hat{s} = \{1 - \hat{\sigma}_u^2(K-2)/(W'P_Z W)\}$.

following theorem justifies its use. Donald and Newey (2001) also present a similar result to justify their selection procedure.[7]

**Theorem 3.** *Assumptions 1-3 are satisfied and* $\hat{o}_\epsilon^2 \to_p \sigma_\epsilon^2$, $\hat{o}_{u\epsilon} \to_p \sigma_{u\epsilon}$ *and* $\hat{\lambda}-\lambda \to_p 0$. *Then,* $\{S(\hat{s}^*) - S(s^*)\}/S(s^*) = o_p(1)$.

# 3 The Shrinkage LIML Estimator

We can extend our idea of the shrinkage TSLS into the limited information maximum likelihood (LIML) estimator. The LIML estimator minimizes $(y-W\delta)'P_x(y-W\delta)/\{(y-W\delta)'(y-W\delta)\}$. The shrinkage LIML estimator $\hat{\delta}_{liml,s}$ is defined as:

$$\begin{aligned}\hat{\delta}_{liml,s} &= \underset{\delta}{\text{argmin}} \left\{ (1-s)\frac{(y-W\delta)'P_X(y-W\delta)}{(y-W\delta)'(y-W\delta)} + s\frac{(y-W\delta)'P_{(X,Z)}(y-W\delta)}{(y-W\delta)'(y-W\delta)} \right\} \\ &= \underset{\delta}{\text{argmin}} \frac{(y-W\delta)'P^s(y-W\delta)}{(y-W\delta)'(y-W\delta)}.\end{aligned}$$

Let $v_i = u_i - \epsilon_i\sigma_{u\epsilon}/\sigma_\epsilon^2$ and define $\Sigma_v = E(v_iv_i')$. The next theorem derives the asymptotic mean square error of the shrinkage LIML estimator. We assume that we have the third-moment condition, $E(\epsilon_i^2 v_i) = 0$, to simplify the formula.

**Theorem 4.** *Assumptions 1-3 are satisfied,* $\Sigma_v \neq 0$, $E(||\eta_i||^5|x_i)$ *and* $E(|\epsilon|^5|x_i)$ *are bounded and* $E(\epsilon_i^2 v_i) = 0$. *Then, under* $sK/N \to_p 0$ *and* $1-s = O_p(sK/N)$ *or* $E(f_iZ_i') = 0$, *we have* $\hat{\delta}_{liml,s} \to_p \delta$, $\sqrt{N}(\hat{\delta}_{liml,s} - \delta) \to_d N(0, \sigma_\epsilon^2\bar{H}^{-1})$ *and*

$$S(s) = H^{-1}\left\{\sigma_\epsilon^2\Sigma_v\frac{s^2K}{N} + \sigma_\epsilon^2\frac{f'(I-P^s)(I-P^s)f}{N}\right\}H^{-1}.$$

The proof of Theorem 4 is in the Appendix. The proof is similar to the proof of Proposition 2 in Donald and Newey (2001). As before, we propose to choose

[7]Note that, in general, this result does not imply that the estimator with $\hat{s}^*$ attains the minimum of $S(s)$. If $\hat{s}^*$ were constructed using samples that are independent of the data used in estimation of $\delta$, then this theorem would imply that the estimators with $\hat{s}^*$ have the same second-order MSE properties as those with $s^*$. However, we usually use the same samples to estimate $s^*$ and $\delta$, which makes it very difficult to prove that the estimators with $\hat{s}^*$ are second-order equivalent to those with $s^*$.

the shrinkage parameter by minimizing $\lambda' S(s)\lambda$ with respect to $s$. The optimal shrinkage parameter is

$$s^* = \frac{\sigma_\epsilon^2 \dfrac{\lambda' H^{-1} f' P_Z f H^{-1} \lambda}{N}}{\lambda' H^{-1} \Sigma_v H^{-1} \lambda \dfrac{K}{N} + \sigma_\epsilon^2 \dfrac{\lambda' H^{-1} f' P_Z f H^{-1} \lambda}{N}}.$$

If there is only one endogenous variable, minimization does not depend on $\lambda$ and the optimal shrinkage parameter has the form:

$$s^* = \frac{\sigma_\epsilon^2 \dfrac{\bar{Y}' P_Z \bar{Y}}{N}}{(\sigma_\eta^2 \sigma_\epsilon^2 - \sigma_{\eta\epsilon}^2) \dfrac{K}{N} + \sigma_\epsilon^2 \dfrac{\bar{Y}' P_Z \bar{Y}}{N}},$$

where $\sigma_\eta^2 = E(\eta_i^2)$ and $\sigma_{\eta\epsilon} = E(\eta_i \epsilon_i)$.

The optimal shrinkage parameter has a $K$-order term, while that of the shrinkage TSLS has a $K^2$-order term. We should shrink the effect of the other instruments less in the shrinkage LIML than in the shrinkage TSLS when $K$ is large. This observation is consistent with the established result that the LIML estimator is more robust against the number of instruments than is the TSLS estimator (see, e.g., Anderson, Kunitomo and Matsushita (2008)). We also note that the optimal shrinkage parameter always lies between 0 and 1 ($0 \leq s^* \leq 1$).

We note that the James-Stein shrinkage parameter has a $K$-order term too. In fact, for both the optimal and the James-Stein shrinkage parameters, the order of $1 - s$ is $K/N$ when $f' P_Z f / N \to_p c > 0$. This observation implies that the James-Stein shrinkage parameter has an optimal property in terms of rate in the LIML estimation. However, the James-Stein shrinkage parameter does not take the correlation between the error term of the first-stage regression and that of the second-stage regression into account and thus the estimator based on the James-Stein shrinkage parameter is not expected to behave well.

# 4 Monte Carlo simulation

This section reports the results of the Monte Carlo experiments.[8] The aims of these experiments are to see how the shrinkage estimators behave in finite samples and to compare the shrinkage methods with other estimation methods. Comparison with the instrument selection procedure in Donald and Newey (2001) is one of the main purposes of this study. To make this comparison easier, we borrow their experimental design.

## 4.1 Design

Our data-generating process is the following model:

$$\begin{aligned} y_i &= W_i\delta + \epsilon_i, \\ W_i &= \pi'(X_i, \bar{Z}_i')' + u_i, \end{aligned}$$

for $i = 1, \ldots, N$, where $W_i$ is a scalar, $\delta$ is the scalar parameter of interest, $X_i$ is a scalar random variable, $\bar{Z}_i$ is a $K \times 1$ vector of random variables, $(X_i, \bar{Z}_i')' \sim i.i.d.N(0, I_{K+1})$ and[9]

$$\begin{pmatrix} \epsilon_i \\ u_i \end{pmatrix} \sim i.i.d.N\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & c \\ c & 1 \end{pmatrix} \right).$$

Let $\bar{K}$ be the total number of instruments so that $\bar{K} = K + 1$. The variable $X_i$ is the "main" instrument, and $\bar{Z}_i$ is the vector of additional instruments. We fix the true value of $\delta$ at $\delta = 0.1$, and we examine how well each estimator estimates $\delta$.

In this framework, each experiment is indexed by the vector of specifications: $(N, \bar{K}, c, \{\pi\})$, where $N$ represents the sample size. We use $N = 100$ and $N = 500$,

[8] This Monte Carlo simulation was conducted with Linux Ox 4.1a (Doornik (2006)).

[9] We also consider a design under which $(\epsilon_i, u_i)$ is generated by the product of a random variable with $t$–distribution with degree of freedom 3 and bivariate normal random variables. The results from this non-normal design are very similar to those reported in the paper and are not presented here. The additional tables that summarize the results from the non-normal design are available from the author upon request.

and we set $\bar{K} = 20$ if $N = 100$ and $\bar{K} = 25$ if $N = 500$. The degree of endogeneity is summarized in $c$, and we set $c = 0.1$, 0.5 and 0.9. The number of replications is 5000 for $N = 100$ and 2500 for $N = 500$.

Hahn and Hausman (2002) observe that the theoretical $R^2$ of the first stage regression is given by $R_f^2 = \pi'\pi/(\pi'\pi+1)$. While we try four different specifications of $\pi$, which are stated later, we specify $\pi$ such that $\pi$ always satisfies $\pi'\pi = R_f^2/(1 - R_f^2)$. We try $R_f^2 = 0.1$ and 0.01.

The first specification of $\pi$ is a case where the instruments are all equally important.

$$\text{Model (a):} \quad \pi_k = \sqrt{\frac{R_f^2}{\bar{K}(1 - R_f^2)}}, \forall k.$$

This case is difficult, as not only are all the instruments equally important, they also are all weak. Using only the first instrument is not appropriate. Using all the instruments might cause the "many-instruments" problem. As there is no reason to prefer some to others, selection methods are not very effective. This is also problematic for shrinkage methods as the main instrument itself is weak and the other instruments are as important as the main one.

The second model considered is

$$\text{Model (b):} \quad \pi_1 = c(\bar{K}), \pi_k = \frac{c(\bar{K})}{\sqrt{\bar{K} - 1}} \forall k > 1,$$

where $c(\bar{K})$ is chosen to satisfy $\pi'\pi = R_f^2/(1 - R_f^2)$. The first instrument is strong but others are weak. This data-generating process seems relevant to many applications. Often, we know that the instruments at hand guarantee the identification of the parameter of interest. However, the estimate using only those instruments has a relatively large standard error, which prevents us from drawing sharp conclusions. In this case, even if we are aware that other possible instruments are relatively weak, we may want to increase the number of instruments to reduce the standard error.

Thirdly, we consider the data-generating process used in Donald and Newey

(2001).

$$\text{Model (c):} \quad \pi_k = c(\bar{K}) \left(1 - \frac{k}{\bar{K}+1}\right)^4.$$

The strength of the instruments decreases moderately in $k$. An instrumental selection procedure, such as that proposed by Donald and Newey (2001), would be suitable in this situation.

Lastly, we consider the following data-generating process:

$$\text{Model (d):} \quad \pi_k = 0 \text{ for } k \leq \frac{\bar{K}}{2}; \quad \pi_k = c(\bar{K}) \left(1 - \frac{k - \bar{K}/2}{\bar{K}/2+1}\right)^4 \text{ for } k > \frac{\bar{K}}{2}.$$

The first half of the instruments are completely redundant and the second half of the instruments are informative. In this sense, the instrument ordering in Model (d) is "wrong" and the results of the design would be useful to see the effect of the pre-specified ordering of instruments in selection methods and the effect of a "wrongly chosen" main instrument in shrinkage methods.

A number of estimators are studied. The first is the TSLS estimator with all available instruments (TSLS). The second is the TSLS estimator with Donald and Newey's (2001) optimal selection of the number of instruments (DNTSLS). The third estimator is the TSLS estimator with Hall and Peixe's (2003) selection of the number of instruments (HP). The next two estimators are the shrinkage TSLS estimators with different choices of shrinkage parameters. The first estimator uses the true optimal shrinkage parameter (OSTSLS), which is infeasible in practice. The performance of OSTSLS can be seen as the upper bound of the performance of shrinkage procedures. The other estimator uses the estimated (i.e., feasible) optimal shrinkage parameter (STSLS). We consider three LIML-type estimators: the LIML estimator with all instruments (LIML), the LIML estimator with the Donald and Newey (2001) optimal selection of the number of instruments (DNLIML), and the (feasible) shrinkage LIML estimator (SLIML). Lastly, we consider the REQML estimator with all available instruments (REQML).

To compute DNTSLS, STSLS, DNLIML and SLIML, preliminary estimates are obtained with the number of instruments that minimizes the first-stage cross-

validation criteria.[10] The cross-validation criteria are used for $\hat{R}(K)$ (see Donald and Newey (2001)) in the selection criteria by Donald and Newey (2001). For the Hall and Peixe (2003) method, we use the penalty function that corresponds to the BIC (Equation (14) in Hall and Peixe (2003)).

The selection methods (DNTSLS, HP and DNLIML) are applied given the ordering of the instruments so that they choose only the number of instruments. The ordering of instruments in each model is associated with the index $k$. We note that the instruments are ordered according to their strength in Models (a)-(c). On the other hand, the order of the instruments is "wrong" in Model (d) in the sense that redundant instruments come first in the ordering. The shrinkage methods are applied with the first instrument as the "main" instrument. We note that the first instrument is the strongest instrument in Models (a)-(c). However, in Model (d), the first instrument is redundant.

## 4.2 Measures

For each estimator, we compute the median bias (median bias), the difference between the 0.1 and 0.9 quantile (Dec. Reg.), the median absolute deviation (MAD), following Donald and Newey (2001). We use these "robust" measures because of concerns about the existence of moments of estimators. For example, it is well known that the LIML estimator does not possess any finite moments and, in fact, we encounter an extremely large value of the mean square error of the LIML estimator in the simulations. A disadvantage of using these robust measures is that the relationship between the theoretical results and the simulation results becomes less clear. To overcome this issue at least partially, we also compute the

[10] Alternatively, we may use the Mallows criteria to choose the number of instruments for preliminary estimates. Using the Mallows criteria reduces the computational time substantially, particularly when the sample size is large, and the results do not change by much. Nonetheless, in this experiment, we use the cross-validation criteria following Donald and Newey (2001).

root-mean truncated square error (RMTSE):

$$\sqrt{E\left[\min\left\{(\hat{\delta}-\delta)^2, 2\right\}\right]}$$

for each estimator $\hat{\delta}$.[11] This measure is always finite and should be closely related to the mean square error.

We also compute the coverage rate (Cov. Rate) of a 95% confidence interval based on each estimator. To construct the confidence intervals to compute the coverage probabilities, we use the following estimate of asymptotic variance, following Donald and Newey (2001). The estimators examined here, except REQML, have the common form: $\hat{\delta} = (\hat{W}'W)^{-1}\hat{W}'y$ (i.e., $\hat{W} = W$ for OLS, $\hat{W} = P_{(X,Z)}W$ for in TSLS etc.). The estimates of variance, $\hat{V}$, are given by:

$$\hat{V} = \frac{1}{N}\hat{\epsilon}'\hat{\epsilon}(\hat{W}'W)^{-1}\hat{W}'\hat{W}(W'\hat{W})^{-1},$$

where $\hat{\epsilon} = y - W\hat{\delta}$. For the REQML estimator, the coverage probability is that of the confidence interval obtained by inverting the likelihood ratio test (see Chamberlain and Imbens (2004, page 302)).

We also compute the coverage probabilities based on Bekker's (1994) asymptotic variance estimator for the LIML-type estimators (see also Hansen, Hausman and Newey (2008)). It is denoted "B. Cov. Rate" in the tables. Bekker's asymptotic variance estimator is consistent for the asymptotic variance even if the number of instruments is proportional to the sample size and has the following formula:

$$\hat{V}_B = \frac{1}{N - trace(P)}\hat{\epsilon}'\hat{\epsilon}(\hat{W}'W)^{-1}\left(\hat{W}'\hat{W} - \lambda\frac{W'(I-P)\hat{\epsilon}\hat{\epsilon}'(I-P)W}{\hat{\epsilon}'(I-P)\hat{\epsilon}}\right)(W'\hat{W})^{-1},$$

where $\lambda = \hat{\epsilon}'P\hat{\epsilon}/(\hat{\epsilon}'\hat{\epsilon})$ and $P$ is $P_{(X,Z)}$ for LIML, the projection matrix spanned by the selected instruments for DNLIML and $P^s$ for SLIML. There are two differences between $\hat{V}$ and $\hat{V}_B$. One is that $\hat{V}$ has $N$ but $\hat{V}_B$ has the total number of degrees of freedom, $N - trace(P)$. It makes $\hat{V}$ tend to be smaller than $\hat{V}_B$. The other is that the matrix in the middle for $\hat{V}$ is $\hat{W}'\hat{W}$ but the corresponding matrix for $\hat{V}_B$

[11] A similar measure is used in Chamberlain and Imbens (2004).

is $\hat{W}'\hat{W}$ minus some positive definite matrix. This difference makes $\hat{V}$ tend to be larger than $\hat{V}_B$. Because of these competing effects, we cannot determine which of $\hat{V}$ and $\hat{V}_B$ is larger in general.

## 4.3 Results

The results of the experiments are summarized in Tables 1–6.[12] The mark '$*$' indicates that the number is more than 1,000.

[Tables 1-6 around here]

First, we summarize the performance of TSLS and LIML. If the endogeneity is small ($c = 0.1$), TSLS performs well. The bias of TSLS is negligible and the diversity of TSLS is also very small. We note that all the estimators have negligible biases when $c = 0.1$, but the "Dec. Reg." of TSLS is smaller than that of other estimators. The coverage rate based on TSLS is also close to 0.95 in those cases. On the other hand, LIML outperforms TSLS in the cases with $c = 0.9$. In those cases, the bias of TSLS is very large and the coverage rate based on TSLS is too low. LIML has a relatively small bias and yields better coverage rates in those cases. Nonetheless, when $c = 0.9$ and $R_f^2 = 0.01$, LIML exhibits some bias and the confidence interval based on LIML is not so reliable.

[12] Another result of the experiments which is interesting and is not included in the tables is about the computational times of the estimators. For illustration, the computational time of each estimator relative to TSLS in Model (a) with $N = 100$ and $R_f^2 = 0.1$ is presented in the following table.

| TSLS | DNTSLS | HP | OSTSLS | STSLS | LIML | DNLIML | SLIML | REQML |
|---|---|---|---|---|---|---|---|---|
| 1 | 74.19 | 6.23 | 3.07 | 42.56 | 21.40 | 60.38 | 74.85 | 160.09 |

We note that the computational times of DNTSLS, STSLS, DNLIML and STSLS include the time required to obtain the preliminary estimate. Unfortunately, the computational times depend heavily on the actual implementation of each procedure (i.e., how to obtain the preliminary estimates and how to estimate $\hat{R}(K)$ for the selection methods of Donald and Newey (2001)), which makes it difficult to provide a conclusive argument.

We now compare selection methods and shrinkage methods. The first comparison is between DNTSLS and HP. While HP typically yields better coverage rates than DNTSLS, HP does not perform as well as DNTSLS in other measures. We note that the method of Hall and Peixe is intended to detect (completely) redundant instruments and it does not take into account the bias-variance trade-off in the number of instruments. However, the instruments in Models (a)-(c) are not completely redundant, although they are weak. The instruments in Model (d) are ordered wrongly and the second half of the instruments are not redundant. This is perhaps the reason why Hall and Peixe's method does not work well in the current setting.

The next comparison is between DNTSLS and STSLS. Generally, STSLS performs well in models (a) and (b), and DNTSLS does well in model (c), although STSLS is better in model (c) with little endogeneity. Typically, the good performance of STSLS is because the diversity of STSLS is less than that of DNTSLS, which is indicated by the values of "Dec. Reg." of STSLS and DNTSLS. A remarkable phenomenon is that there are cases where DNTSLS performs substantially worse than TSLS does. On the other hand, STSLS is usually better than TSLS is. When the endogeneity is small, both DNTSLS and STSLS are outperformed by TSLS, but the performance of STSLS is better compared to that of DNTSLS. A similar phenomenon is observed when we compare DNLIML and SLIML. While SLIML achieves improvement on LIML generally, the relative performance of DNLIML compared with LIML is not stable; in some cases, DNLIML does much better than LIML but there are also cases where DNLIML does much worse than LIML. DNLIML usually performs well in the low-endogeneity cases where TSLS-type estimators perform well. In the high-endogeneity cases that are suitable for LIML-type estimators, SLIML is usually best in models (a) and (b). In model (c) with $c = 0.9$, DNLIML is usually best though the differences between DNLIML and SLIML are small. The selection methods do not work well in Model (d) except that DNLIML improves LIML when the degree of endogeneity is low

($c = 0.1$). On the other hand, the performances of the shrinkage methods are similar to those of the estimators that use all the available instruments. In Model (d), the main instrument is weak and it is optimal to set the shrinkage parameter very close to 1 so that the optimal shrinkage estimators are very similar to the estimators that use all the instruments. The Monte Carlo results show that the shrinkage methods perform as well as they can when the main instrument is weak, while the selection methods may not perform well in such a situation.

REQML exhibits a very small bias, even when the LIML-type estimators exhibit non-negligible biases. The coverage rate of the confidence interval based on REQML is also close to the nominal level in any situation. These results are consistent with the findings of Flores-Lagunes (2007). However, REQML has a very large diversity (i.e., its "Dec. Reg." is very large). Because of the large diversity, REQML is not attractive in terms of "MAD" nor "RMTSE".

Lastly, we compare STSLS with OSTSLS to see the effect of the estimation errors in the shrinkage parameter. We notice that, in some cases, in particular when $c = 0.9$ and $R_f^2 = 0.01$, these two perform differently, although we do also observe cases in which their performances are similar. We may be able to obtain a better estimator by improving the estimation of the shrinkage parameter, but this is beyond the scope of this paper.

The confidence intervals based on the LIML-type estimators are conservative in many cases and using Bekker's method makes the confidence intervals have coverage rates close to 0.95 in those cases. However, when the coverage rate is much smaller than 0.95, using Bekker's method tends to intensify this problem.[13] The confidence intervals based on STSLS are improved by using Bekker's method. However, REQML is still better in terms of coverage rate.

[13]Hansen, Hausman and Newey (2008) report that Bekker's method improves the coverage rates. However, they compare Bekker's asymptotic variance estimator with $(\hat{\epsilon}'\hat{\epsilon}/N)(\hat{W}'W)^{-1}$, not $\hat{V}$.

We conclude this section by summarizing the Monte Carlo results. REQML is highly recommended if we are concerned about bias or coverage rate. If we are concerned about the risk of the estimators (either in terms of median absolute deviation or in terms of mean square error), then we should consider selection methods or shrinkage methods. Selection methods are recommended when the rank ordering of the strength of the instruments is clear. Otherwise, shrinkage methods are recommended. Moreover, we observe that shrinkage methods generally can improve the estimators with all the instruments, while there are cases in which selection methods may perform substantially worse than just using all instruments does.

# 5 Discussion

The idea of shrinkage as stated in this paper can easily be extended into general moment restriction models, although finding an optimal way to shrink the effect of the moment conditions might be demanding. Several extensions are found in Okui (2005), which considers conditional moment restriction models and dynamic panel data models. Investigating a way to choose the shrinkage parameter in general moment restriction models is of interest although it might be challenging. We leave this problem for future investigation. We may also consider a method that chooses $s$ and $K$ simultaneously to minimize the asymptotic mean square error. Such a shrinkage-selection hybrid method may be worth further investigation.

Another useful extension is to handle multiple groups of instruments. Note that this paper focuses on the situation when we have only two groups of instruments: main instruments and others. If we have more than two groups of instruments, we need to shrink them group by group. The optimal shrinkage parameter would be calculated in a way similar to that presented here. The crucial assumption is that we know to which group some particular instrument belongs. We may also think of hybrid methods of adaptively partitioning instruments and shrinking them

group by group. For an estimation of a multivariate normal mean, George (1986) provides an interesting discussion of a method for handling a situation with several candidates for partition. We consider hybrid methods to be a promising direction for future research.

In this paper, we assume that all the instruments are orthogonal to the error term. However, it is also important to examine the validity of instruments in practice. As considered by Hall and Peixe (2003), we may apply the method of Andrews (1999) first in order to eliminate invalid instruments and then apply the shrinkage method. Investigating the properties of such a procedure is also an interesting future research topic.

Finally, higher-order efficiency property of selection and shrinkage estimators would be an interesting theoretical question. Takeuchi and Morimune (1985) shows a higher-order efficiency of LIML-type estimators whose asymptotic bias is adjusted. Anderson, Kunitomo and Matsushita (2008) and Anderson, Kunitomo and Matsushita (2010) obtain a similar result in the presence of many instruments. Their framework excludes the possibility of instrument selection or shrinkage estimation. It is an important future research topic to explore how to discuss efficiency properties of IV estimators with instrument selection and shrinkage estimators.

# A Proofs

This appendix contains the proofs of the theorems. Hereafter, all expectations are conditional on $x$. We follow the same steps as the derivation of the asymptotic mean square error in Donald and Newey (2001). Some of the results used as lemmas are proved in that paper. We will employ Lemma A.1 in Donald and Newey (2001) to show Theorems 2 and 4. The estimator examined has the form $\sqrt{N}(\hat{\delta} - \delta) = \hat{H}^{-1}\hat{h}$. We define $h = f'\epsilon/\sqrt{N}$ and $H = f'f/N$.

**Lemma 1** (Donald and Newey (2001) Lemma A.1)**.** *If there is a decomposition* $\hat{h} = h + T^h + Z^h$ , $\hat{H} = H + T^H + Z^H$,

$$(h + T^h)(h + T^h)' - hh'H^{-1}T^{H\prime} - T^H H^{-1}hh' = \hat{A}(s) + Z^A(s),$$

*such that* $T^h = o_p(1)$, $h = O_p(1)$, $H = O_p(1)$, *the determinant of* $H$ *is bounded away from zero with probability 1,* $\rho_{K,N} = o_p(1)$,

$$||T^H||^2 = o_p(\rho_{K,N}), \quad ||T^h||||T^H|| = o_p(\rho_{K,N}), \quad ||Z^h|| = o_p(\rho_{K,N}), \quad ||Z^H|| = o_p(\rho_{K,N}),$$
$$Z^A(s) = o_p(\rho_{K,N}), \quad E\{\hat{A}(s)|x\} = \sigma^2 H + HS(s)H + o_p(\rho_{K,N}),$$

*then*

$$\begin{aligned} N(\hat{\delta}-\delta_0)(\hat{\delta}-\delta_0)' &= \hat{Q}(s)+\hat{r}(s), \\ E\{\hat{Q}(s)|x\} &= \sigma_\epsilon^2 H^{-1}+S(s)+T(s), \\ \{\hat{r}(s)+T(s)\}/tr(S(s)) &= o_p(1), K\to\infty, N\to\infty. \end{aligned}$$

We state two technical lemmas and their proofs. Those lemmas will be used to prove the theorems. First, recall that $Z'X=0$ and $P^s=P_X+sP_Z$, where $P_X=X(X'X)^{-1}X'$ and $P_Z=Z(Z'Z)^{-1}Z'$.

**Lemma 2.** *Suppose Assumptions 1-3 are satisfied. Then we have 1)* $tr(P^s)=m+sK$*; 2)* $\sum_i (P^s_{ii})^2=o_p(sK)$*; 3)* $\sum_{i\neq j}P^s_{ii}P^s_{jj}=(m+sK)^2+o_p(sK)$*; 4)* $\sum_{i\neq j}P^s_{ij}P^s_{ij}=(m+s^2K)+o_p(sK)$*; 5)* $h=f'\epsilon/\sqrt{N}=O_p(1)$ *and* $H=f'f/N=O_p(1)$*.*

*Proof.* First note that $(sK)^{-1}=O_p(1)$. For part 1,

$$tr(P^s)=tr(P_X)+s\cdot tr(P_Z)=m+sK.$$

Assumption 3 and Lemma 2(1) imply that

$$\sum_i (P^s_{ii})^2\le \max_i(P^s_{ii})tr(P^s)=o_p(1)(m+sK)=o_p(sK).$$

This proves part 2.

Also, these results imply that

$$\sum_{i\neq j}P^s_{ii}P^s_{jj}=\sum_i P^s_{ii}\sum_j P^s_{jj}-\sum_i (P^s_{ii})^2=(m+sK)^2+o_p(sK)$$

which shows part 3.

To show 4, first we observe that

$$\sum_{i\neq j}P^s_{ij}P^s_{ij}=tr(P^{s\prime}P^s)-\sum_i (P^s_{ii})^2.$$

Now, $P^{s\prime}P^s=(P_X+sP_Z)(P_X+sP_Z)=P_X+s^2P_Z$ and $tr(P^{s\prime}P^s)=m+s^2K$. As we know, $\sum_i (P^s_{ii})^2=o_p(sK)$ from part 2 in this lemma,

$$\sum_{i\neq j}P^s_{ij}P^s_{ij}=m+s^2K+o_p(sK).$$

Part 5 is Lemma A.2 (v) in Donald and Newey (2001). □

Let $e^s_f=f'(I-P^s)(I-P^s)f/N$ and $\Delta_s=tr(e^s_f)$.

**Lemma 3.** *Suppose Assumptions 1-3 are satisfied and* $s\to_p 1$ *or* $E(f_iZ_i')=0$*. Then, we have 1)* $\Delta_s=o_p(1)$*; 2)* $f'(I-P^s)\epsilon/\sqrt{N}=O(\Delta_s^{1/2})$*; 3)* $u'P^s\epsilon=O_p(sK)$*; 4)* $E(u'P^s\epsilon\epsilon'P^su)=\sigma_{u\epsilon}\sigma'_{u\epsilon}(m+sK)^2+(\sigma_\epsilon^2\Sigma_u+\sigma_{u\epsilon}\sigma'_{u\epsilon})(m+s^2K)+o_p(sK)$*; 5)* $E(f'\epsilon\epsilon'P^su)=\sum_i f_iP^s_{ii}E(\epsilon_i^2u_i')=O_p(sK)$*; 6)* $\Delta_s^{1/2}/\sqrt{N}=o_p(sK/N+\Delta_s)$*; 7)* $E(hh'H^{-1}u'f/N)=\sum_i f_if_i'H^{-1}E(\epsilon^2u_i)f_i'/(N^2)=O_p(1/N)$*; 8)* $E\{f'(I-P^s)\epsilon\epsilon'P^su/N\}=o_p(\Delta_s^{1/2}\sqrt{sK}/\sqrt{N})$.

*Proof.* As $(I-P^s)(I-P^s) = I - P + (s-1)^2 P_Z$ by simple algebra,

$$\frac{f'(I-P^s)(I-P^s)f}{N} = \frac{f'(I-P)f}{N} + (s-1)^2 \frac{f'P_Z f}{N}.$$

The first term is $o_p(1)$ by Lemma A.3(i) in Donald and Newey (2001) and the second term converges on 0 if $s \to_p 1$ or $f'P_Z f/N \to_p 0$. Therefore, $\Delta_s = o_p(1)$.

Next, we observe that $E\{f'(I-P^s)\epsilon/\sqrt{N}\} = 0$ and

$$E\left\{\frac{f'(I-P^s)\epsilon}{\sqrt{N}}\frac{\epsilon'(I-P^s)f}{\sqrt{N}}\right\} = \sigma_\epsilon^2 \frac{f'(I-P^s)(I-P^s)f}{N} = \sigma_\epsilon^2 e_f^s.$$

Therefore, $f'(I-P^s)\epsilon/\sqrt{N} = O_p(\Delta_s^{1/2})$ by the Chebyshev inequality. This shows 2.

For part 3, the Cauchy-Schwartz inequality says that each element of $u'P^s\epsilon$ is less than $[tr(u'P^s u)(\epsilon'P^s\epsilon)]^{1/2}$. As $E(u'P^s u) = \sigma_u^2(m+sK) = O_p(sK)$ and similarly $E(\epsilon'P^s\epsilon) = O_p(sK)$, the Markov inequality implies that $u'P^s\epsilon/\sqrt{N} = O_p(sK/\sqrt{N})$.

To give 4, observe that $E(u_i P_{ij}^s \epsilon_j \epsilon_k P_{kl}^s u_l') = 0$ if one of $(i,j,k,l)$ is different from all the rest. Also, $E(\epsilon_i^2 u_i u_i')$ is bounded by Assumption 1. Therefore, we have

$$\begin{aligned}
E(u'P^s\epsilon\epsilon'P^s u) &= \sum_i (P_{ii}^s)^2 E(\epsilon_i^2 u_i u_i') + \sum_{i\neq j} E(u_i P_{ii}^s \epsilon_i \epsilon_j P_{jj}^s u_j') \\
&\quad + \sum_{i\neq j} E(u_i P_{ij}^s \epsilon_j \epsilon_i P_{ij}^s u_j') + \sum_{i\neq j} E(u_i P_{ij}^s \epsilon_j^2 P_{ji}^s u_i') \\
&= O_P(1)\sum_i (P_{ii}^s)^2 + \sigma_{u\epsilon}\sigma_{u\epsilon}' \sum_{i\neq j} P_{ii}^s P_{jj}^s + (\sigma_\epsilon \Sigma_u + \sigma_{u\epsilon}\sigma_{u\epsilon}') \sum_{i\neq j} P_{ij}^s P_{ij}^s \\
&= o_p(sK) + \sigma_{u\epsilon}\sigma_{u\epsilon}'(m+sK)^2 + (\sigma_\epsilon^2 \Sigma_u + \sigma_{u\epsilon}\sigma_{u\epsilon}')(m+s^2K),
\end{aligned}$$

by Lemmas 2.2, 2.3 and 2.4.

Assumption 1 also implies that

$$E(f'\epsilon\epsilon'P^s u) = \sum_{i.j.k} f_i P_{jk}^s E(\epsilon_i \epsilon_j u_k') = \sum_i f_i P_{ii}^s E(\epsilon_i^2 u_i')$$

and, furthermore, together with Assumption 3 and Lemma 2.1, that

$$\left|\sum_i f_i P_{ii}^s E(\epsilon_i^2 u_i')\right| \leq \sum_i P_{ii}^s \cdot ||f_i|| \cdot ||E(\epsilon_i^2 u_i')|| = O_p(sK),$$

which gives 5.

To prove 6, first we consider the function of $a$: $sK/a + a$, which is convex and the minimum value of which is $2\sqrt{sK}$ with the minimizer $a = \sqrt{sK}$. If $\Delta_s = 0$, then $(\Delta_s/\sqrt{N})\{(sK)/N + \Delta_s\} = 0$ and for $\Delta_s \neq 0$, $(\Delta_s/\sqrt{N})/\{(sK)/N + \Delta_s\} = (sK/\sqrt{\Delta_s N} + \sqrt{\Delta_s N})^{-1} \leq 1/\sqrt{sK} \to 0$ as $sK \to \infty$.

Part 7 is Lemma A.3(vii) in Donald and Newey (2001).

For part 8, let $Q = I - P^s$ and for some $a$ and $b$ let $\zeta_i = f_a(x_i, z_i)$ and $\mu_i = E(\epsilon_i^2 u_{ib})P_{ii}^s$. Now, the $(a,b)$-th element of $E\{f'(I-P^s)\epsilon\epsilon'P^s u\}$ satisfies

$$\left|E\left(\sum_{i.j.k.l} \zeta_i Q_{ij}\epsilon_j\epsilon_k P_{kl}^s u_{lb}\right)\right| = \left|\sum_{i,j} \zeta_i Q_{ij} E(\epsilon_j^2 u_{jb}) P_{jj}^s\right| = |\zeta' Q\mu| \leq |\zeta' Q Q \zeta|^{1/2} |\mu'\mu|^{1/2},$$

where the inequality is the Cauchy-Schwartz inequality. Now $|\zeta' Q Q \zeta|^{1/2} = O_p((N\Delta_s)^{1/2})$ by the definition of $\Delta_s$. $|\mu'\mu| \leq C\sum_i (P^s_{ii})^2$ for some constant $C$ by Assumption 1 and applying Lemma 2(2) we have $|\mu'\mu| = o_p(sK)$. Therefore, we have

$$E\{f'(I-P^s)\epsilon\epsilon' P^s u/N\} = O_p((N\Delta_s)^{1/2}) o_p(\sqrt{sK}) O_p(1/N) = o_p(\Delta_s^{1/2}\sqrt{sK}/\sqrt{N}).$$

□

## A.1 Proof of Theorems 1 and 2

*Proof.* The shrinkage TSLS estimator has the form:

$$\sqrt{N}(\hat{\delta}_{tsls,s} - \delta_0) = \hat{H^s}^{-1}\hat{h^s}, \quad \hat{H^s} = W' P^s W/N, \quad \hat{h^s} = W' P^s \epsilon/\sqrt{N}.$$

Also, $\hat{H^s}$ and $\hat{h^s}$ are decomposed as

$$\begin{aligned}
\hat{h^s} &= h + T_1^h + T_2^h, \\
T_1^h &= -f'(I-P^s)\epsilon/\sqrt{N}, \quad T_2^h = u' P^s \epsilon/\sqrt{N} \\
\hat{H^s} &= H + T_1^H + T_2^H + Z^H \\
T_1^H &= -f'(I-P^s)f/N, \quad T_2^H = (u'f + f'u)/N \\
Z^H &= \{u' P^s u - u'(I-P^s)f - f'(I-P^s)u\}/N.
\end{aligned}$$

We show that the conditions of Lemma 1 are satisfied and $S(s)$ has the form given in the theorem. Note that it is enough to show that the term is $o_p((sK)^2/N + \Delta_s)$ in order to show that a term is $o_p(\rho_{K,N})$ because $o_p((sK)^2/N + \Delta_s) = o_p(\rho_{K,N})$.

Now, $h = O_p(1)$ and $H = O_p(1)$ by Lemma 2(5). As $T^h = T_1^h + T_2^h = -f'(I - P^s)\epsilon/\sqrt{N} + u' P^s \epsilon/\sqrt{N}$, Lemma 3(2) and 3(3) say that $T_1^h = O_p(\Delta_s^{1/2})$ and $T_2^h = O_p(sK/\sqrt{N})$ so $T^h = O_p(\Delta_s^{1/2}) + O_p(sK/\sqrt{N})$. $\Delta_s = o_p(1)$ by Lemma 3(1) and $sK/\sqrt{N} = o_p(1)$ by $(sK)^2/N = o_p(1)$. Therefore, $T^h = o_p(1)$.

Next,

$$T_1^H = -\frac{f'(I-P^s)f}{N} = -e_f^s - s(1-s)\frac{f' P_Z f}{N} = -e_f^s + O_p\left(\frac{K^2}{N}\right) = O_p\left(\frac{(sK)^2}{N} + \Delta_s\right).$$

$T_2^H = O_p(1/\sqrt{N})$ by the CLT. Note that $(sK)^2/N + \Delta_s = o_p(1)$. Then, each of $\{(sK)^2/N + \Delta_s\}^2$, $N^{-1}$ and $\{(sK)^2/N + \Delta_s\}/N$ are $o(\rho_{K,N})$, which implies $||T^H||^2 = o_p(\rho_{K,N})$.

Now, we analyze $||T^h|| \cdot ||T^H||$. We have seen that $T^h = O_p(\Delta_s^{1/2}) + O_p(sK/\sqrt{N})$. and $T^H = O_p(\Delta_s) + O_p(1/\sqrt{N})$. Now, $O_p(\Delta_s^{3/2}) = o_p(\Delta_s) = o_p(\rho_{K,N})$ by Lemma 3(1), $O_p(\Delta_s^{1/2}/\sqrt{N}) = o_p(sK/N + \Delta_s) = o_p(\rho_{K,N})$ by Lemma 3(6), $O_p(sK\Delta_s/\sqrt{N}) = o_p(\rho_{K,N})$ as $sK\Delta_s^{1/2}/\sqrt{N} \leq (sK)^2/N + \Delta_s$ and $\Delta_s^{1/2} = o_p(1)$ by Lemma 3(1), and $O_p(sK/N) = o_p(\rho_{K,N})$. Therefore, $||T^h|| \cdot ||T^H|| = o_p(\rho_{K,N})$.

As $||Z^h|| = 0$ in our case, $||Z^h|| = o_p(\rho_{K,N})$. The last part, for which we need to show $o_p(\rho_{K,N})$, is $||Z^H||$. Now, $Z^H = u' P^s u/N - u'(I-P^s)f/N - f'(I-P^s)u/N$, where the first term is $O_p(sK/N) = o_p(\rho_{K,N})$ and the second and third terms are $O_p(\Delta^{1/2}/\sqrt{N}) = o_p(sK/N + \Delta_s) = o_p(\rho_{K,N})$ by Lemma 3(6). Therefore, we have $||Z^H|| = o_p(\rho_{K,N})$.

Note that we have shown $\hat{H} = H + o_p(1)$ and $\hat{h} = h + o_p(1)$. Then, Proposition 1 holds by the LLN, the CLT and the Slutzky Lemma.

The discussion above indicates $Z^A(s) = 0$ and $\hat{A}(s) = (h + T_1^h + T_2^h)(h + T_1^h + T_2^h)' - hh'H^{-1}(T_1^H + T_2^H)' - (T_1^H + T_2^H)H^{-1}hh'$.

Now, we calculate the expectation of each term in $A(s)$. First of all, $E(hh') = E(f\epsilon\epsilon' f'/N) = \sigma_\epsilon^2 H$. Second,

$$E(hT_1^{h\prime}) = E\left\{-\frac{f'\epsilon\epsilon'(I-P^s)f}{N}\right\} = -\sigma_\epsilon^2\frac{f'(I-P^s)f}{N}.$$

Similarly, $E(T_1^h h') = \sigma_\epsilon^2 f'(I-P^s)f/N$. Third,

$$E(hT_2^{h\prime}) = E\left(\frac{f'\epsilon\epsilon' P^s u}{N}\right) = O_p\left(\frac{sK}{N}\right),$$

by Lemma 3(5). This implies that $E(T_2^h h') = O_p(sK/N)$ too. Fourth,

$$E(T_1^h T_1^{h\prime}) = E\left\{\frac{f'(I-P^s)\epsilon\epsilon'(I-P^s)f}{N}\right\} = \sigma_\epsilon^2\frac{f'(I-P^s)(I-P^s)f}{N}.$$

Fifth,

$$E(T_1^h T_2^{h\prime}) = -E\left\{\frac{f'(I-P^s)\epsilon\epsilon' P^s u}{N}\right\} = o_p\left(\frac{\Delta_s^{1/2}\sqrt{sK}}{\sqrt{N}}\right)$$

by Lemma 3(8). Again, we have $E(T_2^h T_1^{h\prime}) = o_p(\Delta_s^{1/2}\sqrt{sK}/\sqrt{N})$. Sixth,

$$E(T_2^h T_2^{h\prime}) = E\left(\frac{u'P^s\epsilon\epsilon' P^s u}{N}\right) = \sigma_{u\epsilon}\sigma_{u\epsilon}'\frac{(sK)^2}{N} + o_p\left(\frac{(sK)^2}{N}\right),$$

by Lemma 3(4). Seventh,

$$E(hh'H^{-1}T_1^{H\prime}) = -E\left\{\frac{f'\epsilon\epsilon' f H^{-1}f'(I-P^s)f}{N^2}\right\} = -\sigma_\epsilon^2\frac{f'(I-P^s)f}{N}.$$

Also, we have $E(T_1^H H^{-1}hh') = -\sigma_\epsilon^2 f'(I-P^s)f/N$. Finally, Lemma 3(7) implies that

$$E(hh'H^{-1}T_2^{H\prime}) = E\left\{\frac{hh'H^{-1}(u'f + f'u)}{N}\right\} = O_p\left(\frac{1}{N}\right)$$

and $E(T_2^H H^{-1}hh') = O_P(1/N)$. Therefore, we have

$$\begin{aligned}
E(\hat{A}(K)) &= \sigma_\epsilon^2 H - \sigma_\epsilon^2\frac{f'(I-P^s)f}{N} + O_p\left(\frac{sK}{N}\right) \\
&\quad -\sigma_\epsilon^2\frac{f'(I-P^s)f}{N} + \sigma_\epsilon^2\frac{f'(I-P^s)(I-P^s)f}{N} + o_p\left(\frac{\Delta_s^{1/2}\sqrt{sK}}{\sqrt{N}}\right) \\
&\quad +O_p\left(\frac{sK}{N}\right) + o_p\left(\frac{\Delta_s^{1/2}\sqrt{sK}}{\sqrt{N}}\right) + \sigma_{u\epsilon}\sigma_{u\epsilon}'\frac{(sK)^2}{N} + o_p\left(\frac{sK}{N}\right) \\
&\quad +\sigma_\epsilon^2\frac{f'(I-P^s)f}{N} + O_p\left(\frac{1}{N}\right) + \sigma_\epsilon^2\frac{f'(I-P^s)f}{N} + O_p\left(\frac{1}{N}\right) \\
&= \sigma_\epsilon^2 H + \sigma_{u\epsilon}\sigma_{u\epsilon}'\frac{(sK)^2}{N} + \sigma_\epsilon^2\frac{f'(I-P^s)(I-P^s)f}{N} + o_p(\rho_{K,N}),
\end{aligned}$$

where the last equality holds because $1/N = o_p(\rho_{K,N})$, $sK/N = o_p(\rho_{K,N})$ and $o_p(\Delta_s^{1/2}\sqrt{sK}/\sqrt{N}) = o_p(\rho_{K,N})$ by the fact that $\Delta_s^{1/2}\sqrt{sK}/\sqrt{N} \leq sK/N + \Delta_s$. □

## A.2 Proof of Theorem 3

*Proof.* Under the assumptions, we have $W'(P_X + P_Z)W/N - H \to_p 0$ and $W'P_ZW/N - f'P_Zf/N \to_p 0$. Let

$$\begin{aligned} V &\equiv \frac{\sigma_\epsilon^2}{\lambda' H^{-1}\sigma_{u\epsilon}\sigma_{u\epsilon}' H^{-1}\lambda}\frac{\lambda' H^{-1}f'P_Zf H^{-1}\lambda}{N}, \\ \hat{V} &\equiv \frac{\hat{\sigma}_\epsilon^2}{\hat{\lambda}'\hat{H}^{-1}\hat{\sigma}_{u\epsilon}\hat{\sigma}_{u\epsilon}'\hat{H}^{-1}\hat{\lambda}}\frac{\hat{\lambda}'\hat{H}^{-1}W'P_ZW\hat{H}^{-1}\hat{\lambda}}{N}. \end{aligned}$$

Then, $\hat{V} - V = o_p(1)$ and $s^*$ and $\hat{s}^*$ can be written as $1 - (1 + VN/K^2)^{-1}$ and $1 - (1 + \hat{V}N/K^2)^{-1}$ respectively. Suppose that $f'P_Zf/N \to c > 0$ for some $c$. Then

$$\hat{s}^* - s^* = \frac{\hat{V}N/K^2 - VN/K^2}{(1 + \hat{V}N/K^2)(1 + VN/K^2)} = \frac{\hat{V} - V}{(K^2/N + \hat{V})(K^2/N + V)}\frac{K^2}{N} = o_p(K^2/N).$$

This implies that

$$\begin{aligned} H\{S(\hat{s}^*) - S(s^*)\}H &= (\hat{s}^{*2} - s^{*2})\sigma_{u\epsilon}\sigma_{u\epsilon}'K^2/N + \{(1 - \hat{s}^*)^2 - (1 - s^*)^2\}\sigma_\epsilon^2 f'P_Zf/N \\ &= o_p(K^2/N), \end{aligned}$$

by the continuous mapping theorem. As $S(s^*)$ is at least $O_p(K^2/N)$ in this case, the result holds.

Suppose that $f'P_Zf = O_p(K)$, which occurs when $Z$ is a matrix of irrelevant instruments. Then, $s^* = O_p(1/K)$ and $S(s^*) = O_p(1/N)$. As $N(\hat{V} - V)/K = o_p(1)$, we have

$$\hat{s}^* - s^* = \frac{\hat{V}N/K^2 - VN/K^2}{(1 + \hat{V}N/K^2)(1 + VN/K^2)} = \frac{N(\hat{V} - V)/K}{(1 + \hat{V}N/K^2)(1 + VN/K^2)}\frac{1}{K} = o_p(1/K).$$

It follows therefore that $S(\hat{s}^*) - S(s^*) = o_p(1/N)$.

□

## A.3 Proof of Theorem 4

First, we show that the consistency of the shrinkage LIML and derive the asymptotic distribution of it under $sK/N \to 0$. Now, our $\hat{\delta}$ is $\hat{\delta}_{liml,s} = \text{argmin}_\delta (y - W\delta)'P^s(y - W\delta)/(y - W\delta)'(y - W\delta)$.

**Lemma 4.** *Suppose that Assumptions 1-3 are satisfied. Then, under $sK/N \to 0$ and $s \to 1$ or $E(f_iZ_i') = 0$, it follows that $\hat{\delta}_{liml,s} \to_p \delta_0$.*

*Proof.* Define $\bar{W} \equiv (y, W)$ and $D_0 \equiv (\delta, I)$. $\bar{W}$ can be written as $\bar{W} = WD_0 + \epsilon e_1$, where $e_1$ is the first unit vector. Let $\hat{A} = \bar{W}'P^s\bar{W}/N$ and $A = D_0'\bar{H}D_0$.

Observing Lemma A.4 and the proof of Lemma A.5 in Donald and Newey (2001), it is enough to show that $\hat{A} \to_p A$.

The term $\hat{A}$ has the following decomposition:

$$\begin{aligned}\hat{A} &= D_0' \left\{ \frac{f'f}{N} - \frac{f'(I-P^s)f}{N} + \frac{u'P^sf}{N} + \frac{f'P^su}{N} \right\} D_0 \\ &\quad + e_1 \frac{\epsilon' P^s W}{N} D_0 + D_0' \frac{W P^s \epsilon}{N} e_1' + \frac{\epsilon' P^s \epsilon}{N} e_1 e_1'.\end{aligned}$$

First, we have $f'f/N \to_p \bar{H}$ by the LLN. $f'(I-P^s)f/N = f'(I-P)f/N + (1-s)f'P_Zf/N \to_p 0$ by Lemma A.2(1) in Donald and Newey (2001) and where $s \to_p 1$ or $f'P_Zf/N \to_p 0$. $E(\epsilon' P^s \epsilon) = tr\{P^s E(\epsilon\epsilon')\} = \sigma_\epsilon^2(m+sK)$, which implies that $\epsilon' P^s \epsilon/N \to_p 0$ by the Markov inequality. Similarly, we can show that $u'P^su/N \to_p 0$. Let $W_j$ be the $j$th column of $W$. Then,

$$\left| \frac{W_j' P^s \epsilon}{N} \right| \leq \sqrt{\frac{W_j' P^s W_j}{N}} \sqrt{\frac{\epsilon' P^s \epsilon}{N}} \leq \sqrt{\frac{W_j' W_j}{N}} o_p(1) = o_p(1).$$

The first inequality is the Cauchy-Schwartz inequality and the second inequality comes from the fact that $I-P^s$ is positive definite, which is because $I-P^s = I-P+(1-s)P_Z$ and $I-P$ and $P_Z$ are positive definite and $1-s \geq 0$. It follows therefore that $W'P^s\epsilon/N \to_p 0$. $f'P^su/N \to_p 0$ similarly. Summing up, we have $\hat{A} \to_p A$. □

**Lemma 5.** *Suppose that Assumptions 1-3 are satisfied, $sK/N \to_p 0$ and $s \to_p 1$ or $E(f_iZ_i') = 0$. Then, we have $\sqrt{N}(\hat{\delta}_{liml,s} - \delta_0) \to_d N(0, \sigma_\epsilon^2 \bar{H}^{-1})$.*

*Proof.* Let $A^s(\delta) \equiv (y - W\delta)' P^s (y - W\delta)/N$ and $B(\delta) \equiv (y-W\delta)'(y-W\delta)/N$. Define $\Lambda(\delta) \equiv A^s(\delta)/B(\delta)$ so that $\hat{\delta}_{liml,s} = \text{argmin}_\delta \Lambda(\delta)$.

Let $\Lambda_\delta(\delta)$ and $\Lambda_{\delta\delta}(\delta)$ be the gradient and Hessian of $\Lambda(\delta)$, respectively. A standard Taylor expansion shows that

$$\sqrt{N}(\hat{\delta}_{liml,s} - \delta_0) = -\Lambda_{\delta\delta}(\tilde{\delta})^{-1} \sqrt{N} \Lambda_\delta(\delta_0) = \left\{ \frac{\tilde{\sigma}_\epsilon^2 \Lambda_{\delta\delta}(\tilde{\delta})}{2} \right\}^{-1} \left\{ -\frac{\tilde{\sigma}_\epsilon^2 \sqrt{N} \Lambda_\delta(\delta_0)}{2} \right\},$$

for some mean value $\tilde{\delta}$. Now, we have

$$\begin{aligned}\Lambda_\delta(\delta) &= B(\delta)^{-1}\{A_\delta(\delta) - \Lambda(\delta) B_\delta(\delta)\}, \\ \Lambda_{\delta\delta}(\delta) &= B(\delta)^{-1}\{A_{\delta\delta}(\delta) - \Lambda(\delta) B_{\delta\delta}(\delta)\} - B(\delta)^{-1}\{B_\delta(\delta)\Lambda_\delta(\delta)' + \Lambda_\delta(\delta) B_\delta'\}.\end{aligned}$$

As $\hat{\delta}_{liml,s} \to_p \delta_0$ by Lemma 4, $\tilde{\delta} \to_p \delta_0$, which implies that $B(\tilde{\delta}) \to_p \sigma_\epsilon^2$, $B_\delta(\tilde{\delta}) \to_p -2\sigma_{u\epsilon}$. As before, $A(\tilde{\delta}) \to_p 0$, which implies that $\Lambda(\tilde{\delta}) \to_p 0$. Also, we have $A_\delta(\tilde{\delta}) \to_p 0$, which gives $\Lambda_\delta(\tilde{\delta}) \to_p 0$. Lastly, we have $A_{\delta\delta}(\delta) = 2W'P^sW/N \to_p 2\bar{H}$ and $B_{\delta\delta}(\delta) = 2W'W/N \to_p 2E(W_iW_i')$. Therefore, we have $\tilde{\sigma}_\epsilon^2 \Lambda_{\delta\delta}(\tilde{\delta}) \to_p \bar{H}$.

Consider the gradient term. First, define $\hat{\alpha} = W'\epsilon/\epsilon'\epsilon$ and $\alpha = \sigma_{u\epsilon}/\sigma_\epsilon^2$. $\hat{\alpha} - \alpha = O_p(1/N)$ by the CLT. We have the following decomposition:

$$-\frac{\tilde{\sigma}_\epsilon^2 \sqrt{N} \Lambda_\delta(\delta_0)}{2} = \frac{W'P^s\epsilon}{\sqrt{N}} - \frac{\epsilon' P^s \epsilon W'\epsilon}{\sqrt{N}\epsilon'\epsilon} = h - \frac{f'(I-P^s)\epsilon}{\sqrt{N}} + \frac{v'P^s\epsilon}{\sqrt{N}} - (\hat{\alpha} - \alpha)\frac{\epsilon' P^s \epsilon}{\sqrt{N}}.$$

$h \to_d N(0, \sigma^2 \bar{H})$ by the CLT. Lemma 2(1) and the Chebyshev inequality says $f'(I - P^s)\epsilon/\sqrt{N} = o_p(1)$. A similar argument as in the proof of Lemma 2(4) together with

$E(v_i\epsilon_i)=0$ implies that $v'P^s\epsilon/\sqrt{N}=O_p(\sqrt{sK/N})=o_p(1)$. $\epsilon'P^s\epsilon=O_p(sK)$ as we see in the proof of Lemma 4. It follows, therefore, that $(\hat{\alpha}-\alpha)\epsilon'P^s\epsilon/\sqrt{N}=O_p(sK/N)=o_p(1)$. We have $-\tilde{\sigma}_\epsilon^2\sqrt{N}\Lambda_\delta(\delta_0)/2\to_d N(0,\sigma_\epsilon^2\bar{H})$.

In conclusion, we have $\sqrt{N}(\hat{\delta}_{liml,s}-\delta)\to_d \bar{H}^{-1}N(0,\sigma_\epsilon^2\bar{H})=N(0,\sigma_\epsilon^2\bar{H}^{-1})$. □

Define $\hat{\Lambda}=\min_\delta(y-W\delta)'P^s(y-W\delta)/(y-W\delta)'(y-W\delta)$ and $\tilde{\Lambda}=\epsilon'P^s\epsilon/(N\sigma_\epsilon^2)$. Also note that in the LIML case, to show that a term is $o_p(\rho_{K,N})$, it is enough to show that it is $o_p(sK/N+\Delta_s)$.

**Lemma 6.** *Suppose Assumptions 1-3 are satisfied, $sK/N\to_p 0$ and $1-s=o_p(K/N)$ or $E(f_iZ_i')=0$. Then, it follows that*

$$\hat{\Lambda}=\tilde{\Lambda}-\left(\frac{\tilde{\sigma}_\epsilon^2}{\sigma_\epsilon^2}-1\right)\tilde{\Lambda}-\frac{h'H^{-1}h}{2N\sigma_\epsilon^2}+\hat{R}_\Lambda=\tilde{\Lambda}+o_p\left(\frac{sK}{N}\right)$$

*and $\sqrt{N}\hat{R}_\Lambda=o_p(\rho_{K,N})$.*

*Proof.* We expand $\hat{\Lambda}=\Lambda(\hat{\delta})$ around the true value $\delta_0$. Then

$$\begin{aligned}\hat{\Lambda} &= \Lambda(\delta_0)-\frac{\Lambda_\delta(\delta_0)'\{\Lambda_{\delta\delta}(\delta_0)\}^{-1}\Lambda_\delta(\delta_0)}{2}+O_p\left(\frac{1}{N^{3/2}}\right)\\ &= \tilde{\Lambda}-\left(\frac{\tilde{\sigma}_\epsilon^2}{\sigma_\epsilon^2}-1\right)\tilde{\Lambda}+\frac{(\tilde{\sigma}_\epsilon^2-\sigma_\epsilon^2)^2}{\tilde{\sigma}_\epsilon^2\sigma_\epsilon^2}\tilde{\Lambda}-\frac{\Lambda_\delta(\delta_0)'\{\Lambda_{\delta\delta}(\delta_0)\}^{-1}\Lambda_\delta(\delta_0)}{2}+O_p\left(\frac{1}{N^{3/2}}\right).\end{aligned}$$

We can see from the proof of Lemma 5 that

$$-\frac{\tilde{\sigma}_\epsilon^2\sqrt{N}\Lambda_\delta(\delta_0)}{2}=h+O_p\left(\Delta_s^{1/2}+\frac{sK}{N}\right).$$

This also implies that $\Lambda_\delta=O_p(1/\sqrt{N})$. Then,

$$\frac{\tilde{\sigma}_\epsilon^2\Lambda_{\delta\delta}(\tilde{\delta})}{2}=\frac{W'P^sW}{N}-\Lambda(\delta_0)\frac{W'W}{N}+O_p\left(\frac{1}{\sqrt{N}}\right)$$

by $B_\delta(\delta_0)=O_p(1)$. It follows that

$$\frac{\tilde{\sigma}_\epsilon^2\Lambda_{\delta\delta}(\tilde{\delta})}{2}=H-\frac{f'(I-P^s)f}{N}+\frac{u'P^sf}{N}+\frac{f'P^su}{N}+\frac{u'P^su}{N}+O_p\left(\sqrt{\frac{sK}{N}}\right)$$

by $\Lambda(\delta_0)=O_p(\sqrt{sK/N})$. As in the proof of proposition 2, we have $f'(I-P^s)f/N=O_p(\Delta_s+sK/N)$. It holds also that $u'P^sf/N=O_p(1/\sqrt{N})$ and $u'P^su/N=O_p(sK/N)$. Summing up, we have $\tilde{\sigma}_\epsilon^2\Lambda_{\delta\delta}(\tilde{\delta})/2=H+O_p(\Delta_s^{1/2}+\sqrt{sK/N})$. Then, we have

$$\frac{\Lambda_\delta(\delta_0)'\{\Lambda_{\delta\delta}(\delta_0)\}^{-1}\Lambda_\delta(\delta_0)}{2}=\frac{h'H^{-1}h}{N\sigma_\epsilon^2}+O_p\left(\frac{\Delta_s^{1/2}}{N}+\sqrt{\frac{sK}{N^3}}\right).$$

Also, it follows that $\left(\tilde{\sigma}_\epsilon^2/\sigma_\epsilon^2-1\right)=O_p(1/\sqrt{N})$ by the CLT and the Delta method. These results give the first equation of the lemma as

$$\hat{\Lambda} = \tilde{\Lambda}-\left(\frac{\tilde{\sigma}_\epsilon^2}{\sigma_\epsilon^2}-1\right)\tilde{\Lambda}-\frac{h'H^{-1}h}{2N\sigma_\epsilon^2}+O_p\left(\frac{\Delta_s^{1/2}}{N}+\sqrt{\frac{sK}{N^3}}\right)+O_p\left(\frac{sK}{N^2}\right)+O_p\left(\frac{1}{N^{3/2}}\right),$$

and all the remainder terms are $o_p(\rho_{K,N})$.

The second equation in the lemma is given by the fact that $\tilde{\Lambda}=O_p(sK/N)$. □

**Lemma 7.** *Suppose that Assumptions 1-3 are satisfied, $sK/N \to_p 0$. Then,*

1. $u'P^s u/N - \tilde{\Lambda}\Sigma_u = o_p(sK/N)$,
2. $E(h\tilde{\Lambda}\epsilon' v/\sqrt{N}) = (m+sK)/N \cdot \sum_i f_i E(\epsilon_i^2 v_i')/N + O_p(sK/N^2)$,
3. $E(hh'H^{-1}h/\sqrt{N}) = O_p(1/N)$.

*Proof.* We begin with the proof of 1. $E(\tilde{\Lambda}) = tr\{P^s E(\epsilon\epsilon')\}/(N\sigma_\epsilon^2) = (m+sK)/N$ and

$$
\begin{aligned}
E\left\{\left(\tilde{\Lambda} - \frac{m+sK}{N}\right)^2\right\} &= \frac{E(\epsilon' P^s \epsilon \epsilon' P^s \epsilon)}{N^2\sigma_\epsilon^4} - \left(\frac{m+sK}{N}\right)^2 \\
&= \frac{\sigma_\epsilon^4(m+sK)^2 + o_p((sK)^2)}{N^2\sigma_\epsilon^4} - \left(\frac{m+sK}{N}\right)^2 = o_p\left(\left(\frac{sK}{N}\right)^2\right),
\end{aligned}
$$

by Lemma 3(4) by replacing $u$ with $\epsilon$. This gives $\{\tilde{\Lambda} - (m+sK)/N\}\Sigma_u = o_p(sK/N)$. We also have $E(u'P^s u) = (m+sK)\Sigma_u$ and $u'P^s u/N - \{(m+sK)/N\}\Sigma_u = o_p(sK/N)$. Therefore, 1 is proved.

We observe that

$$
\begin{aligned}
E\left(\frac{h\tilde{\Lambda}\epsilon' v}{\sqrt{N}}\right) &= \frac{\sum_{i,j,k,l} E(f_i \epsilon_i \epsilon_j P^s_{jk} \epsilon_k \epsilon_l v_l')}{N^2\sigma_\epsilon^2} \\
&= \frac{\sum_i f_i P^s_{ii} E(\epsilon_i^4 v_i')}{N^2\sigma_\epsilon^2} + 2\frac{\sum_{i\neq j} f_i P^s_{ij} E(\epsilon_j^2 v_j')}{N^2} + \frac{\sum_{i\neq j} f_i P^s_{jj} E(\epsilon_i^2 v_i')}{N^2} \\
&= O_p\left(\frac{sK}{N^2}\right) + o_p\left(\frac{sK}{N^2}\right) + \frac{m+sK}{N}\frac{\sum_i f_i E(\epsilon_i^2 v_i')}{N},
\end{aligned}
$$

which gives 2.

Part 3 is Lemma A.8(iii) in Donald and Newey (2001). □

*Proof of Theorem 4.* The consistency and the asymptotic normality of the shrinkage estimator stems from Lemmas 4 and 5. The shrinkage LIML estimator has the following representation:

$$
\sqrt{N}(\hat{\delta}_{liml,s} - \delta_0) = \hat{H}^{-1}\hat{h}, \quad \hat{H} = \frac{W'P^s W}{N} - \hat{\Lambda}\frac{W'W}{N}, \quad \hat{h} = \frac{W'P^s\epsilon}{\sqrt{N}} - \hat{\Lambda}\frac{W'\epsilon}{\sqrt{N}}.
$$

As in the case of TSLS, we verify the assumption of Lemma 1. First, $\hat{H}$ and $\hat{h}$ have the following decomposition:

$$
\begin{aligned}
\hat{h} &= h + \sum_{i=1}^5 T_i^h + Z^h, \\
T_1^h &= -\frac{f'(I-P^s)\epsilon}{\sqrt{N}} = O_p(\Delta_s^{1/2}), \quad T_2^h = \frac{v'P^s\epsilon}{\sqrt{N}} = O_p\left(\sqrt{\frac{sK}{N}}\right), \\
T_3^h &= -\tilde{\Lambda}h = O_p\left(\frac{sK}{N}\right), \quad T_4^h = -\tilde{\Lambda}\frac{v'\epsilon}{\sqrt{N}} = O_p\left(\frac{sK}{N}\right), \\
T_5^h &= \frac{h'H^{-1}h}{2\sqrt{N}\sigma_\epsilon^2}\sigma_{u\epsilon} = O_p\left(\frac{1}{\sqrt{N}}\right), \\
Z^h &= -(\hat{\Lambda}-\tilde{\Lambda})h + \sqrt{N}\left(\frac{\tilde{\sigma}_\epsilon^2}{\sigma_\epsilon^2} - 1\right)\tilde{\Lambda}\left(\frac{u'\epsilon}{N} - \sigma_{u\epsilon}\right) + \frac{h'H^{-1}h}{2\sqrt{N}\sigma_\epsilon^2}\left(\frac{u'\epsilon}{N} - \sigma_{u\epsilon}\right) - \sqrt{N}\hat{R}_\Lambda\frac{u'\epsilon}{N},
\end{aligned}
$$

and

$$\begin{aligned}
\hat{H} &= H + \sum_{i=1}^{3} T_i^H + Z^H, \\
T_1^H &= -\frac{f'(I-P^s)f}{N} = O_p\left(\frac{sK}{N} + \Delta_s\right), \quad T_2^H = \frac{u'f + f'u}{N} = O_p\left(\frac{1}{\sqrt{N}}\right), \\
T_3^H &= -\tilde{\Lambda}H = O_p\left(\frac{sK}{N}\right), \\
Z^H &= -\frac{u'(I-P^s)f}{N} - \frac{f'(I-P^s)u}{N} + \frac{uP^su}{N} \\
&\quad -\tilde{\Lambda}\frac{u'u}{N} - \tilde{\Lambda}\left(\frac{u'f + f'u}{N}\right) - (\hat{\Lambda} - \tilde{\Lambda})\frac{W'W}{N}.
\end{aligned}$$

$h = O_p(1)$ and $H = O_p(1)$ by Lemma 3(8). $T^h = o_p(1)$ as all of $\Delta_s^{1/2}$, $\sqrt{sK/N}$, $sK/N$ and $1/\sqrt{N}$ are $o_p(1)$.

$||T_1^H||^2$ consists of terms of order $(sK/N + \Delta_s)^2$, $1/N$, $(sK/N)^2$, $(sK/N + \Delta_s)/\sqrt{N}$, $(sK/N + \Delta_s) \cdot sK/N$ and $sK/N^{3/2}$. It is easy to see that all of them are $o_p(\rho_{K,N})$. It follows that $||T_1^H||^2 = o_p(\rho_{K,N})$.

Similarly, $||T^h|| \cdot ||T^H||$ consists of terms of order $(sK/N + \Delta_s)o_p(1)$, $\Delta_s^{1/2}/\sqrt{N}$, $\sqrt{sK}/N$, $1/N$ and $sK/N \cdot o_p(1)$. A simple inspection and Lemma 3(6) say that all of them are $o_p(\rho_{K,N})$. That gives $||T^h|| \cdot ||T^H|| = o_p(\rho_{K,N})$.

To show $Z^h = o_p(\rho_{K,N})$, we investigate each term of $Z^h$. $(\hat{\Lambda} - \tilde{\Lambda})h = o_p(sK/N)O_p(1) = o_p(\rho_{K,N})$ by Lemma 6. $\sqrt{N}(\tilde{\sigma}_\epsilon^2/\sigma_\epsilon^2 - 1)\tilde{\Lambda}(u'\epsilon/N - \sigma_{u\epsilon}) = O_p(1)O_p(sK/N)O_p(1/\sqrt{N}) = O_p(sK/N^{3/2}) = o_p(\rho_{K,N})$ by the CLT and the delta method. $h'H^{-1}h/(2\sqrt{N}\sigma_\epsilon^2) \cdot (u'\epsilon/N - \sigma_{u\epsilon}) = O_p(1/\sqrt{N})O_p(1/\sqrt{N}) = O_p(1/N) = o_p(\rho_{K,N})$ by the CLT. $\sqrt{N}\hat{R}_\Lambda u'\epsilon/N = o_p(\rho_{K,N})O_p(1) = o_p(\rho_{K,N})$ by the LLN and Lemma 6. Therefore, $Z^h = o_p(\rho_{K,N})$.

Similarly, each term of $Z^H$ is shown to be $o_p(\rho_{K,N})$. $u'(I - P^s)f/N = O_p(\Delta_s^{1/2}/\sqrt{N}) = o_p(\rho_{K,N})$ where the first equality can be verified as in the proof of Lemma 3(2) and the second equality is Lemma 3(6). $uP^su/N - \tilde{\Lambda}u'u/N = uP^su/N - \tilde{\Lambda}\Sigma_u - \tilde{\Lambda}(u'u/N - \Sigma_u) = o_p(sK/N) + O_p(sK/N)o_p(1) = o_p(\rho_{K,N})$ by Lemma 7(1) and the LLN. The CLT implies that $\tilde{\Lambda}(u'f + f'u)/N = O_p(sK/N)O_p(1/\sqrt{N}) = o_p(\rho_{K,N})$. Finally $(\hat{\Lambda} - \tilde{\Lambda})W'W/N = o_p(sK/N)O_p(1) = o_p(\rho_{K,N})$ by the LLN and Lemma 6. Hence, we have $Z^H = o_p(\rho_{K,N})$.

Consider the decomposition:

$$\left(h + \sum_{i=1}^{5} T_i^h\right)\left(h + \sum_{i=1}^{5} T_i^h\right)' - hh'H^{-1}\sum_{i=1}^{3} T_i^{H\prime} - \sum_{i=1}^{3} T_i^H H^{-1}hh' = A(s) + Z^A(s),$$

where

$$\begin{aligned}
A(s) &\equiv hh' + \sum_{i=1}^{5} hT_i^{h\prime} + \sum_{i=1}^{5} T_i^{h\prime}h' + (T_1^h + T_2^h)(T_1^h + T_2^h)' \\
&\quad -hh'H^{-1}\sum_{i=1}^{3} T_i^{H\prime} - \sum_{i=1}^{3} T_i^H H^{-1}hh', \\
Z^A(s) &\equiv \left(\sum_{i=3}^{5} T_i^h\right)\left(\sum_{i=3}^{5} T_i^h\right)' + \left(\sum_{i=3}^{5} T_i^h\right)(T_1^h + T_2^h)' + (T_1^h + T_2^h)\left(\sum_{i=3}^{5} T_i^h\right)'.
\end{aligned}$$

$Z^A(s)$ consists of terms of order $(sK/N)^2$, $sK/N^{3/2}$, $1/N$, $\Delta_s^{1/2}sK/N$, $\Delta_s^{1/2}/\sqrt{N}$, $(sK/N)^{3/2}$ and $\sqrt{sK}/N$. All of them are $o_p(\rho_{K,N})$ by a simple inspection and Lemma 3(6). $Z^A(s) = o_p(\rho_{K,N})$.

What remains to be shown is the expectation of $A(s)$. As we saw in the TSLS case, we have $E(hh') = \sigma_\epsilon^2 H$, $E(hT_1^{h\prime}) = E(T_1^h h') = -\sigma_\epsilon^2 f'(I-P^s)f/N$, $E(T_1^h T_1^{h\prime}) = \sigma_\epsilon^2 f'(I-P^s)(I-P^s)f/N$, $E(T_1^h T_2^{h\prime}) = o_p(\Delta_s^{1/2}\sqrt{sK/N}) = o_p(\rho_{K,N})$, similarly $E(T_2^h T_1^{h\prime}) = o_p(\rho_{K,N})$, $E(hh'H^{-1}T_1^{H\prime}) = E(T_1^H H^{-1}hh') = -\sigma_\epsilon^2 f'(I-P^s)f/N$, $E(hh'H^{-1}T_2^{H\prime}) = O_p(1/N) = o_p(\rho_{K,N})$ and similarly $E(T_2^H H^{-1}hh') = o_p(\rho_{K,N})$.

A similar argument as in the proof of Lemma 3(6), noting that $E(v_i\epsilon_i) = 0$, gives

$$E(T_2^h T_2^{h\prime}) = \sigma_\epsilon^2 \Sigma_v \frac{s^2 K}{N} + o_p(\rho_{K,N}).$$

Lemma 7(3) shows

$$E(hT_5^{h\prime}) = E\left(\frac{hh'H^{-1}h}{2N\sigma_\epsilon^2}\sigma_{u\epsilon}\right) = O_p\left(\frac{1}{N}\right) = o_p(\rho_{K,N}).$$

Similarly, $E(T_5^h h) = o_p(\rho_{K,N})$.

Lemma 7(2) gives

$$E(hT_4^{h\prime}) = E\left(\frac{h\tilde{\Lambda}\epsilon' v}{\sqrt{N}}\right) = -\frac{sK}{N}\frac{\sum_i f_i E(\epsilon_i^2 v_i')}{N} + o_p(\rho_{K,N}).$$

Also, we have $E(hT_2^{h\prime}) = \sum_i f_i P_{ii}^s E(\epsilon_i^2 v_i')/N$. Letting $\hat{\zeta} \equiv \sum_i f_i P_{ii}^s E(\epsilon_i^2 v_i')/N - sK/N \cdot \sum_i f_i E(\epsilon_i^2 v_i')/N$, $E(hT_2^{h\prime}) + E(hT_4^{h\prime}) = \hat{\zeta} + o_p(\rho_{K,N})$ and $E(T_2^h h') + E(T_4^h h') = \hat{\zeta}' + o_p(\rho_{K,N})$.

Summing up, we have

$$\begin{aligned} E(A(s)) &= \sigma_\epsilon^2 H - 2\sigma_\epsilon^2 \frac{f'(I-P^s)f}{N} + \hat{\zeta} + \hat{\zeta}' \\ &\quad + \sigma_\epsilon^2 \frac{f'(I-P^s)(I-P^s)f}{N} + \sigma_\epsilon^2 \Sigma_v \frac{s^2 K}{N} + 2\sigma_\epsilon^2 \frac{f'(I-P^s)f}{N} + o_p(\rho_{K,N}) \\ &= \sigma_\epsilon^2 H + \sigma_\epsilon^2 \Sigma_v \frac{s^2 K}{N} + \sigma_\epsilon^2 \frac{f'(I-P^s)(I-P^s)f}{N} + \hat{\zeta} + \hat{\zeta}' + o_p(\rho_{K,N}). \end{aligned}$$

Note that under $E(\epsilon_i^2 v_i) = 0$, $\hat{\zeta} = 0$. □

# References

Andrews, D. W. K. (1999). Consistent moment selection procedures for generalized method of moments estimation, *Econometrica* **67**(3): 543–564.

Anderson, T. W., Kunitomo, N. and Matsushita, Y. (2008). On finite sample properties of alternative estimators of coefficients in a structural equation with many instruments, unpublished manuscript.

Anderson, T. W., Kunitomo, N. and Matsushita, Y. (2010). On the asymptotic optimality of the LIML estimator with possibly many instruments, *Journal of Econometrics* **157**: 191-204.

Angrist, J. D. and Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings?, *Quarterly Journal of Economics* **106**(4): 979–1014.

Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators, *Econometrica* **62**(3): 657–681.

Bound, J., Jaeger, D. A. and Baker, R. M. (1996). Problems with instrumental variables estimation when correlation between the instruments and the endogenous explanatory variable is weak, *Journal of the American Statistical Association* **90**(430): 443–450.

Carrasco, M. (2008). A regularization approach to the many instruments problem, unpublished manuscript.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions, *Journal of Econometrics* **34**: 305–334.

Chamberlain, G. and Imbens, G. (2004). Random effects estimators with many instrumental variables, *Econometrica* **72**(1): 295–306.

Chao, J. C. and Swanson, N. R. (2005). Consistent estimation with a large number of weak instruments, *Econometrica* **73**(5): 1673–1692.

Donald, S. G. and Newey, W. K. (2001). Choosing the number of instruments, *Econometrica* **69**(5): 1161–1191.

Doornik, J. A. (2006). *An Object-oriented Matrix Programming Language - Ox 4*, Timberlake Consultants Ltd.

Flores-Lagunes, A. (2007). Finite sample evidence of IV estimators under weak instruments, *Journal of Applied Econometrics* **22**(3): 677–694.

George, E. I. (1986). Combining minimax shrinkage estimators, *Journal of the American Statistical Association* **81**(394): 437–445.

Hahn, J. and Hausman, J. (2002). A new specification test for the validity of instrumental variables, *Econometrica* **70**(1): 163–189.

Hall, A. R. and Peixe, F. P. M. (2003). A consistent method for the selection of relevant instruments, *Econometric Reviews* **22**(3): 269–287.

Hansen, C., Hausman, J. and Newey, W. K. (2008). Estimation with many instrumental variables, *the Journal of Business and Economic Statistics* **26**(4): 398–422.

Hastie, T., Tibshirani, R. and Friedman, J. (2001). *The Elements of Statistical Learning; data mining, inference, and prediction*, Springer, New York.

Kuersteiner, G. M. (2002). Mean squared error reduction for GMM estimators of linear time series models, unpublished manuscript.

Kunitomo, N. (1980). Asymptotic expansions of the distributions of estimators in a linear functional relationship and simultaneous equations, *Journal of the American Statistical Association* **75**: 693–700.

Morimune, K. (1983). Approximate distributions of $k$-class estimators when the degree of overidentification is large compared with the sample size, *Econometrica* **51**(3): 821–841.

Morimune, K. (1985). *Keizai Moderu no Suitei to Kentei (Estimation and Testing in Economic Models)*, (in Japanese), Kyouritsu Shuppan, Tokyo, Japan.

Nagar, A. L. (1959). The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations, *Econometrica* **27**(4): 575–595.

Okui, R. (2005). *Instrumental Variable Estimation with Many Moment Conditions with Applications to Dynamic Panel Data Models*, PhD thesis, University of Pennsylvania.

Phillips, P. C. B. (1983). Exact small sample theory in the simultaneous equations model, *in* Z. Griliches and M. D. Intriligator (eds), *Handbook of Econometrics*, Vol. 1, North-Holland Publishing Company, chapter 8.

Small, D. (2002). *Inference and model selection for instrumental variables regression*, PhD thesis, Stanford University.

Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments, *Econometrica* **65**(3): 557–586.

Stock, J. H. and Yogo, M. (2005). Asymptotic distribution of instrumental variables statistics with many weak instruments, *in* D. W. K. Andrews and J. H. Stock (eds), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Cambridge University Press, pp. 109–120.

Takeuchi, K. and and Morimune, K. (1985). Third-order efficiency of the extended maximum likelihood estimators in a simultaneous equation system, *Econometrica* **53**(1): 177–200.

West, K., Wong, K. and Anatolyev, S. (2009). Instrumental variables estimation of heteroskedastic linear models using all lags of instruments, *Econometric Reviews* **26**(5): 441–467.

Table 1: Monte Carlo results

| model(a) | $R_f^2 = 0.1$ | TSLS | DNTSLS | HP | OSTSLS | STSLS | LIML | DNLIML | SLIML | REQML |
|---|---|---|---|---|---|---|---|---|---|---|
| $c = 0.1$ | | | | | | | | | | |
| $N = 100$ | | | | | | | | | | |
| | median bias | 0.0597 | 0.0521 | 0.0627 | 0.0607 | 0.0608 | 0.037 | 0.058 | 0.0345 | -0.0316 |
| | Dec. Reg. | 0.467 | 1.84 | 3.89 | 0.467 | 0.582 | 1.82 | 1.23 | 1.75 | 3.83 |
| | MAD | 0.129 | 0.275 | 0.643 | 0.129 | 0.149 | 0.347 | 0.283 | 0.341 | 0.478 |
| | RMTSE | 0.195 | 0.67 | 0.905 | 0.196 | 0.272 | 0.67 | 0.537 | 0.656 | 0.829 |
| | Cov. Rate | 0.944 | 0.98 | 0.997 | 0.945 | 0.947 | 0.985 | 0.985 | 0.985 | 0.923 |
| | B. Cov. Rate | | | | | | 0.965 | 0.977 | 0.996 | |
| $N = 500$ | | | | | | | | | | |
| | median bias | 0.0324 | 0.033 | 0.0382 | 0.0321 | 0.0316 | 0.00465 | 0.0109 | 0.00417 | 0.00187 |
| | Dec. Reg. | 0.287 | 0.3 | 1.93 | 0.288 | 0.287 | 0.43 | 0.426 | 0.432 | 0.459 |
| | MAD | 0.0787 | 0.0823 | 0.361 | 0.079 | 0.0786 | 0.111 | 0.11 | 0.111 | 0.116 |
| | RMTSE | 0.117 | 0.152 | 0.696 | 0.117 | 0.118 | 0.175 | 0.172 | 0.176 | 0.301 |
| | Cov. Rate | 0.945 | 0.954 | 0.99 | 0.945 | 0.944 | 0.966 | 0.966 | 0.965 | 0.953 |
| | B. Cov. Rate | | | | | | 0.965 | 0.967 | 0.983 | |
| $c = 0.5$ | | | | | | | | | | |
| $N = 100$ | | | | | | | | | | |
| | median bias | 0.317 | 0.293 | 0.28 | 0.305 | 0.309 | 0.0208 | 0.241 | 0.0315 | -0.0163 |
| | Dec. Reg. | 0.425 | 1.69 | 3.77 | 0.624 | 0.541 | 1.6 | 1.15 | 1.52 | 7.36 |
| | MAD | 0.317 | 0.398 | 0.654 | 0.312 | 0.317 | 0.319 | 0.336 | 0.32 | 0.435 |
| | RMTSE | 0.358 | 0.701 | 0.905 | 0.4 | 0.389 | 0.636 | 0.559 | 0.624 | 0.789 |
| | Cov. Rate | 0.497 | 0.76 | 0.952 | 0.695 | 0.546 | 0.944 | 0.899 | 0.944 | 0.929 |
| | B. Cov. Rate | | | | | | 0.951 | 0.907 | 0.976 | |
| $N = 500$ | | | | | | | | | | |
| | median bias | 0.158 | 0.192 | 0.152 | 0.152 | 0.156 | 0.0044 | 0.0296 | 0.00453 | 0.00528 |
| | Dec. Reg. | 0.265 | 0.423 | 1.88 | 0.289 | 0.268 | 0.411 | 0.408 | 0.411 | 0.432 |
| | MAD | 0.158 | 0.212 | 0.374 | 0.154 | 0.157 | 0.105 | 0.109 | 0.106 | 0.108 |
| | RMTSE | 0.184 | 0.302 | 0.691 | 0.188 | 0.186 | 0.167 | 0.164 | 0.167 | 0.278 |
| | Cov. Rate | 0.666 | 0.69 | 0.932 | 0.716 | 0.672 | 0.961 | 0.948 | 0.961 | 0.954 |
| | B. Cov. Rate | | | | | | 0.962 | 0.949 | 0.976 | |
| $c = 0.9$ | | | | | | | | | | |
| $N = 100$ | | | | | | | | | | |
| | median bias | 0.571 | 0.543 | 0.501 | 0.533 | 0.557 | -0.00869 | 0.32 | 0.00297 | 0.0335 |
| | Dec. Reg. | 0.294 | 1.83 | 3.3 | 0.754 | 0.409 | 1.18 | 1.01 | 1.15 | 1.25 |
| | MAD | 0.571 | 0.61 | 0.677 | 0.534 | 0.557 | 0.234 | 0.383 | 0.232 | 0.254 |
| | RMTSE | 0.582 | 0.792 | 0.904 | 0.623 | 0.58 | 0.497 | 0.561 | 0.488 | 0.595 |
| | Cov. Rate | 0.0152 | 0.551 | 0.765 | 0.426 | 0.147 | 0.94 | 0.722 | 0.937 | 0.951 |
| | B. Cov. Rate | | | | | | 0.927 | 0.692 | 0.951 | |
| $N = 500$ | | | | | | | | | | |
| | median bias | 0.275 | 0.292 | 0.238 | 0.257 | 0.271 | 0.00377 | 0.0318 | 0.00416 | 0.0489 |
| | Dec. Reg. | 0.204 | 1.07 | 1.82 | 0.298 | 0.247 | 0.364 | 0.345 | 0.363 | * |
| | MAD | 0.275 | 0.36 | 0.41 | 0.257 | 0.271 | 0.0937 | 0.097 | 0.0939 | 0.124 |
| | RMTSE | 0.283 | 0.567 | 0.685 | 0.284 | 0.284 | 0.149 | 0.143 | 0.148 | 0.664 |
| | Cov. Rate | 0.176 | 0.693 | 0.774 | 0.468 | 0.311 | 0.959 | 0.94 | 0.959 | 0.964 |
| | B. Cov. Rate | | | | | | 0.95 | 0.929 | 0.955 | |

Table 2: Monte Carlo results

| model(a) | $R_f^2 = 0.01$ | TSLS | DNTSLS | HP | OSTSLS | STSLS | LIML | DNLIML | SLIML | REQML |
|---|---|---|---|---|---|---|---|---|---|---|
| $c = 0.1$ | | | | | | | | | | |
| $N = 100$ | | | | | | | | | | |
| | median bias | 0.0954 | 0.0878 | 0.096 | 0.0948 | 0.0934 | 0.293 | 0.144 | 0.269 | -0.0903 |
| | Dec. Reg. | 0.57 | 4.03 | 5.77 | 0.677 | 0.871 | 16.1 | 2.45 | 11.4 | 31 |
| | MAD | 0.163 | 0.57 | 0.927 | 0.183 | 0.211 | 0.851 | 0.477 | 0.806 | 0.953 |
| | RMTSE | 0.244 | 0.887 | 1.01 | 0.281 | 0.397 | 0.996 | 0.767 | 0.977 | 1.03 |
| | Cov. Rate | 0.944 | 0.991 | 1 | 0.953 | 0.959 | 0.99 | 0.992 | 0.99 | 0.91 |
| | B. Cov. Rate | | | | | | 0.91 | 0.958 | 0.996 | |
| $N = 500$ | | | | | | | | | | |
| | median bias | 0.0872 | 0.0729 | 0.102 | 0.0877 | 0.0881 | 0.133 | 0.13 | 0.125 | -0.0501 |
| | Dec. Reg. | 0.489 | 2.95 | 5.29 | 0.488 | 0.688 | 4.39 | 1.61 | 4 | 8.38 |
| | MAD | 0.141 | 0.446 | 0.853 | 0.144 | 0.175 | 0.607 | 0.386 | 0.579 | 0.721 |
| | RMTSE | 0.208 | 0.807 | 0.989 | 0.21 | 0.318 | 0.875 | 0.639 | 0.861 | 0.945 |
| | Cov. Rate | 0.938 | 0.988 | 1 | 0.941 | 0.954 | 0.988 | 0.986 | 0.989 | 0.916 |
| | B. Cov. Rate | | | | | | 0.936 | 0.962 | 0.994 | |
| $c = 0.5$ | | | | | | | | | | |
| $N = 100$ | | | | | | | | | | |
| | median bias | 0.475 | 0.475 | 0.491 | 0.486 | 0.474 | 0.289 | 0.461 | 0.309 | -0.0581 |
| | Dec. Reg. | 0.514 | 3.48 | 4.87 | 1.85 | 0.763 | 5.51 | 2.02 | 4.87 | * |
| | MAD | 0.475 | 0.703 | 0.93 | 0.599 | 0.481 | 0.868 | 0.603 | 0.834 | 0.983 |
| | RMTSE | 0.514 | 0.916 | 1.02 | 0.797 | 0.582 | 0.995 | 0.821 | 0.981 | 1.03 |
| | Cov. Rate | 0.323 | 0.83 | 0.966 | 0.943 | 0.457 | 0.919 | 0.901 | 0.919 | 0.917 |
| | B. Cov. Rate | | | | | | 0.887 | 0.89 | 0.959 | |
| $N = 500$ | | | | | | | | | | |
| | median bias | 0.422 | 0.38 | 0.389 | 0.411 | 0.416 | 0.112 | 0.379 | 0.13 | -0.0331 |
| | Dec. Reg. | 0.427 | 2.66 | 4.66 | 0.983 | 0.611 | 3.32 | 1.42 | 3.24 | * |
| | MAD | 0.422 | 0.56 | 0.87 | 0.434 | 0.421 | 0.566 | 0.462 | 0.55 | 0.727 |
| | RMTSE | 0.451 | 0.836 | 0.992 | 0.573 | 0.5 | 0.848 | 0.685 | 0.838 | 0.939 |
| | Cov. Rate | 0.276 | 0.776 | 0.976 | 0.79 | 0.406 | 0.926 | 0.902 | 0.924 | 0.926 |
| | B. Cov. Rate | | | | | | 0.904 | 0.89 | 0.959 | |
| $c = 0.9$ | | | | | | | | | | |
| $N = 100$ | | | | | | | | | | |
| | median bias | 0.858 | 0.843 | 0.845 | 0.845 | 0.854 | 0.139 | 0.715 | 0.207 | 0.128 |
| | Dec. Reg. | 0.275 | 1.96 | 2.69 | 1.5 | 0.431 | * | * | * | * |
| | MAD | 0.858 | 0.885 | 0.935 | 0.855 | 0.854 | 0.831 | 0.844 | 0.814 | 0.981 |
| | RMTSE | 0.863 | 0.981 | 1.02 | 0.941 | 0.872 | 0.978 | 0.923 | 0.968 | 1.05 |
| | Cov. Rate | 0.0004 | 0.51 | 0.709 | 0.671 | 0.052 | 0.764 | 0.521 | 0.748 | 0.938 |
| | B. Cov. Rate | | | | | | 0.533 | 0.394 | 0.783 | |
| $N = 500$ | | | | | | | | | | |
| | median bias | 0.749 | 0.723 | 0.707 | 0.722 | 0.739 | 0.0139 | 0.516 | 0.0279 | 0.107 |
| | Dec. Reg. | 0.254 | 2.18 | 3.34 | 1.08 | 0.413 | 4.95 | 1.7 | 4.39 | * |
| | MAD | 0.749 | 0.773 | 0.863 | 0.722 | 0.739 | 0.399 | 0.613 | 0.397 | 0.498 |
| | RMTSE | 0.755 | 0.921 | 0.994 | 0.827 | 0.759 | 0.72 | 0.745 | 0.711 | 0.863 |
| | Cov. Rate | 0.0004 | 0.528 | 0.76 | 0.433 | 0.0668 | 0.896 | 0.591 | 0.894 | 0.956 |
| | B. Cov. Rate | | | | | | 0.812 | 0.505 | 0.898 | |

Table 3: Monte Carlo results

| model(b) | $R_f^2 = 0.1$ | TSLS | DNTSLS | HP | OSTSLS | STSLS | LIML | DNLIML | SLIML | REQML |
|---|---|---|---|---|---|---|---|---|---|---|
| $c = 0.1$ | | | | | | | | | | |
| $N = 100$ | | | | | | | | | | |
| | median bias | 0.0631 | 0.0266 | 0.0119 | 0.0555 | 0.0463 | 0.0351 | 0.031 | 0.0231 | -0.00607 |
| | Dec. Reg. | 0.467 | 0.982 | 1.24 | 0.483 | 0.607 | 1.81 | 0.956 | 1.34 | 1.29 |
| | MAD | 0.13 | 0.224 | 0.282 | 0.13 | 0.153 | 0.354 | 0.233 | 0.292 | 0.288 |
| | RMTSE | 0.196 | 0.453 | 0.536 | 0.199 | 0.269 | 0.668 | 0.419 | 0.578 | 0.554 |
| | Cov. Rate | 0.945 | 0.972 | 0.988 | 0.948 | 0.947 | 0.985 | 0.983 | 0.984 | 0.931 |
| | B. Cov. Rate | | | | | | 0.963 | 0.982 | 0.996 | |
| $N = 500$ | | | | | | | | | | |
| | median bias | 0.0334 | 0.0255 | 0.00701 | 0.0309 | 0.0292 | 0.00478 | 0.0179 | 0.00129 | 0.00276 |
| | Dec. Reg. | 0.285 | 0.345 | 0.482 | 0.288 | 0.305 | 0.431 | 0.406 | 0.405 | 0.412 |
| | MAD | 0.0784 | 0.0892 | 0.127 | 0.079 | 0.0822 | 0.111 | 0.109 | 0.105 | 0.107 |
| | RMTSE | 0.117 | 0.145 | 0.196 | 0.117 | 0.124 | 0.176 | 0.162 | 0.162 | 0.165 |
| | Cov. Rate | 0.944 | 0.946 | 0.968 | 0.946 | 0.944 | 0.962 | 0.956 | 0.96 | 0.944 |
| | B. Cov. Rate | | | | | | 0.962 | 0.96 | 0.978 | |
| $c = 0.5$ | | | | | | | | | | |
| $N = 100$ | | | | | | | | | | |
| | median bias | 0.318 | 0.153 | 0.0405 | 0.104 | 0.221 | 0.021 | 0.137 | 0.0346 | -0.00279 |
| | Dec. Reg. | 0.428 | 1.05 | 1.29 | 0.805 | 0.678 | 1.61 | 0.953 | 1.27 | 1.32 |
| | MAD | 0.318 | 0.286 | 0.285 | 0.224 | 0.264 | 0.323 | 0.259 | 0.281 | 0.28 |
| | RMTSE | 0.357 | 0.481 | 0.54 | 0.337 | 0.341 | 0.635 | 0.448 | 0.554 | 0.544 |
| | Cov. Rate | 0.493 | 0.8 | 0.944 | 0.919 | 0.671 | 0.948 | 0.913 | 0.944 | 0.935 |
| | B. Cov. Rate | | | | | | 0.952 | 0.916 | 0.973 | |
| $N = 500$ | | | | | | | | | | |
| | median bias | 0.155 | 0.0891 | 0.0175 | 0.0604 | 0.104 | 0.00235 | 0.0504 | 0.00415 | 0.00327 |
| | Dec. Reg. | 0.267 | 0.459 | 0.491 | 0.362 | 0.359 | 0.41 | 0.396 | 0.394 | 0.399 |
| | MAD | 0.156 | 0.145 | 0.129 | 0.108 | 0.128 | 0.106 | 0.113 | 0.102 | 0.102 |
| | RMTSE | 0.184 | 0.196 | 0.201 | 0.152 | 0.169 | 0.165 | 0.161 | 0.157 | 0.16 |
| | Cov. Rate | 0.669 | 0.814 | 0.948 | 0.925 | 0.787 | 0.959 | 0.93 | 0.958 | 0.947 |
| | B. Cov. Rate | | | | | | 0.958 | 0.933 | 0.97 | |
| $c = 0.9$ | | | | | | | | | | |
| $N = 100$ | | | | | | | | | | |
| | median bias | 0.571 | 0.155 | 0.0576 | 0.0776 | 0.277 | -0.00256 | 0.158 | 0.0135 | 0.00611 |
| | Dec. Reg. | 0.295 | 1.33 | 1.44 | 0.924 | 0.682 | 1.17 | 1.02 | 1.08 | 1.18 |
| | MAD | 0.571 | 0.332 | 0.293 | 0.248 | 0.305 | 0.228 | 0.271 | 0.22 | 0.238 |
| | RMTSE | 0.581 | 0.546 | 0.565 | 0.408 | 0.379 | 0.494 | 0.466 | 0.469 | 0.554 |
| | Cov. Rate | 0.015 | 0.75 | 0.86 | 0.879 | 0.619 | 0.943 | 0.837 | 0.933 | 0.954 |
| | B. Cov. Rate | | | | | | 0.93 | 0.819 | 0.943 | |
| $N = 500$ | | | | | | | | | | |
| | median bias | 0.276 | 0.0518 | 0.0258 | 0.0433 | 0.0912 | 0.00129 | 0.045 | 0.00458 | 0.0256 |
| | Dec. Reg. | 0.205 | 0.488 | 0.519 | 0.411 | 0.367 | 0.371 | 0.347 | 0.366 | 0.507 |
| | MAD | 0.276 | 0.137 | 0.134 | 0.116 | 0.125 | 0.0935 | 0.101 | 0.0936 | 0.107 |
| | RMTSE | 0.283 | 0.206 | 0.215 | 0.168 | 0.165 | 0.147 | 0.142 | 0.146 | 0.464 |
| | Cov. Rate | 0.168 | 0.88 | 0.911 | 0.919 | 0.842 | 0.959 | 0.928 | 0.956 | 0.954 |
| | B. Cov. Rate | | | | | | 0.95 | 0.918 | 0.954 | |

Table 4: Monte Carlo results

| model(b) | $R_f^2 = 0.01$ | TSLS | DNTSLS | HP | OSTSLS | STSLS | LIML | DNLIML | SLIML | REQML |
|---|---|---|---|---|---|---|---|---|---|---|
| $c = 0.1$ | | | | | | | | | | |
| $N = 100$ | | | | | | | | | | |
| | median bias | 0.0928 | 0.0794 | 0.0774 | 0.0833 | 0.0867 | 0.304 | 0.108 | 0.244 | -0.0327 |
| | Dec. Reg. | 0.568 | 3.29 | 4.45 | 0.816 | 0.884 | 17.1 | 2.11 | 10.6 | 6.09 |
| | MAD | 0.163 | 0.522 | 0.742 | 0.214 | 0.215 | 0.845 | 0.443 | 0.779 | 0.802 |
| | RMTSE | 0.244 | 0.841 | 0.948 | 0.337 | 0.408 | 0.994 | 0.736 | 0.964 | 0.974 |
| | Cov. Rate | 0.944 | 0.99 | 0.999 | 0.967 | 0.96 | 0.99 | 0.994 | 0.991 | 0.908 |
| | B. Cov. Rate | | | | | | 0.908 | 0.965 | 0.997 | |
| $N = 500$ | | | | | | | | | | |
| | median bias | 0.0832 | 0.0396 | 0.0363 | 0.0762 | 0.0715 | 0.141 | 0.0682 | 0.113 | -0.0128 |
| | Dec. Reg. | 0.482 | 1.53 | 2.11 | 0.533 | 0.721 | 4.57 | 1.33 | 3.22 | 2.3 |
| | MAD | 0.141 | 0.328 | 0.431 | 0.147 | 0.184 | 0.593 | 0.325 | 0.499 | 0.451 |
| | RMTSE | 0.207 | 0.628 | 0.734 | 0.222 | 0.324 | 0.875 | 0.566 | 0.81 | 0.755 |
| | Cov. Rate | 0.939 | 0.984 | 0.995 | 0.951 | 0.959 | 0.987 | 0.991 | 0.989 | 0.921 |
| | B. Cov. Rate | | | | | | 0.93 | 0.979 | 0.995 | |
| $c = 0.5$ | | | | | | | | | | |
| $N = 100$ | | | | | | | | | | |
| | median bias | 0.474 | 0.384 | 0.314 | 0.343 | 0.441 | 0.294 | 0.39 | 0.286 | -0.00876 |
| | Dec. Reg. | 0.507 | 3.02 | 4.12 | 2.22 | 0.801 | 5.29 | 1.99 | 4.65 | * |
| | MAD | 0.475 | 0.636 | 0.755 | 0.586 | 0.458 | 0.864 | 0.563 | 0.805 | 0.803 |
| | RMTSE | 0.514 | 0.868 | 0.947 | 0.809 | 0.552 | 0.992 | 0.786 | 0.966 | 0.969 |
| | Cov. Rate | 0.321 | 0.822 | 0.961 | 0.956 | 0.493 | 0.917 | 0.9 | 0.915 | 0.915 |
| | B. Cov. Rate | | | | | | 0.885 | 0.89 | 0.958 | |
| $N = 500$ | | | | | | | | | | |
| | median bias | 0.42 | 0.209 | 0.0844 | 0.158 | 0.329 | 0.124 | 0.224 | 0.121 | -0.00506 |
| | Dec. Reg. | 0.423 | 1.61 | 2.16 | 1.24 | 0.767 | 3.27 | 1.3 | 2.72 | 3.26 |
| | MAD | 0.42 | 0.404 | 0.429 | 0.334 | 0.357 | 0.563 | 0.373 | 0.486 | 0.463 |
| | RMTSE | 0.451 | 0.654 | 0.733 | 0.514 | 0.434 | 0.85 | 0.587 | 0.787 | 0.758 |
| | Cov. Rate | 0.28 | 0.79 | 0.959 | 0.932 | 0.546 | 0.927 | 0.906 | 0.927 | 0.928 |
| | B. Cov. Rate | | | | | | 0.905 | 0.899 | 0.956 | |
| $c = 0.9$ | | | | | | | | | | |
| $N = 100$ | | | | | | | | | | |
| | median bias | 0.855 | 0.683 | 0.518 | 0.514 | 0.787 | 0.147 | 0.556 | 0.215 | 0.0291 |
| | Dec. Reg. | 0.275 | 2.75 | 3.81 | 2.63 | 0.656 | * | * | * | * |
| | MAD | 0.855 | 0.79 | 0.74 | 0.649 | 0.787 | 0.826 | 0.747 | 0.772 | 0.828 |
| | RMTSE | 0.862 | 0.916 | 0.936 | 0.871 | 0.812 | 0.977 | 0.884 | 0.952 | 1 |
| | Cov. Rate | 0.0004 | 0.582 | 0.77 | 0.784 | 0.165 | 0.763 | 0.602 | 0.753 | 0.942 |
| | B. Cov. Rate | | | | | | 0.531 | 0.47 | 0.786 | |
| $N = 500$ | | | | | | | | | | |
| | median bias | 0.749 | 0.258 | 0.106 | 0.143 | 0.444 | 0.00316 | 0.258 | 0.0461 | 0.0569 |
| | Dec. Reg. | 0.258 | 1.81 | 2.15 | 1.34 | 0.743 | 5.03 | 2.17 | 3.4 | * |
| | MAD | 0.749 | 0.488 | 0.4 | 0.352 | 0.451 | 0.396 | 0.428 | 0.374 | 0.584 |
| | RMTSE | 0.755 | 0.698 | 0.714 | 0.589 | 0.536 | 0.718 | 0.651 | 0.683 | 0.942 |
| | Cov. Rate | 0 | 0.697 | 0.871 | 0.859 | 0.465 | 0.897 | 0.764 | 0.895 | 0.953 |
| | B. Cov. Rate | | | | | | 0.81 | 0.695 | 0.892 | |

Table 5: Monte Carlo results

| model(c) | $R_f^2 = 0.1$ | TSLS | DNTSLS | HP | OSTSLS | STSLS | LIML | DNLIML | SLIML | REQML |
|---|---|---|---|---|---|---|---|---|---|---|
| $c = 0.1$ | | | | | | | | | | |
| $N = 100$ | | | | | | | | | | |
| | median bias | 0.061 | 0.0253 | 0.0238 | 0.0582 | 0.0533 | 0.0313 | 0.0215 | 0.0261 | -0.0161 |
| | Dec. Reg. | 0.461 | 0.783 | 1.22 | 0.464 | 0.542 | 1.8 | 0.872 | 1.46 | 1.54 |
| | MAD | 0.128 | 0.189 | 0.26 | 0.127 | 0.14 | 0.35 | 0.214 | 0.307 | 0.32 |
| | RMTSE | 0.194 | 0.39 | 0.545 | 0.196 | 0.249 | 0.669 | 0.394 | 0.606 | 0.623 |
| | Cov. Rate | 0.95 | 0.969 | 0.985 | 0.949 | 0.953 | 0.985 | 0.98 | 0.983 | 0.924 |
| | B. Cov. Rate | | | | | | 0.965 | 0.979 | 0.996 | |
| $N = 500$ | | | | | | | | | | |
| | median bias | 0.0336 | 0.0177 | 0.00834 | 0.0325 | 0.0309 | 0.00508 | 0.00463 | 0.00488 | 0.00458 |
| | Dec. Reg. | 0.288 | 0.33 | 0.389 | 0.29 | 0.297 | 0.432 | 0.365 | 0.421 | 0.423 |
| | MAD | 0.078 | 0.0837 | 0.0997 | 0.0779 | 0.0797 | 0.108 | 0.0937 | 0.106 | 0.107 |
| | RMTSE | 0.117 | 0.131 | 0.16 | 0.117 | 0.12 | 0.177 | 0.145 | 0.169 | 0.171 |
| | Cov. Rate | 0.947 | 0.95 | 0.955 | 0.948 | 0.946 | 0.962 | 0.959 | 0.962 | 0.948 |
| | B. Cov. Rate | | | | | | 0.964 | 0.959 | 0.98 | |
| $c = 0.5$ | | | | | | | | | | |
| $N = 100$ | | | | | | | | | | |
| | median bias | 0.316 | 0.161 | 0.113 | 0.158 | 0.261 | 0.015 | 0.115 | 0.0263 | -0.00307 |
| | Dec. Reg. | 0.431 | 0.798 | 1.21 | 0.763 | 0.587 | 1.6 | 0.829 | 1.33 | 1.61 |
| | MAD | 0.317 | 0.24 | 0.283 | 0.232 | 0.28 | 0.317 | 0.227 | 0.29 | 0.303 |
| | RMTSE | 0.358 | 0.417 | 0.542 | 0.344 | 0.349 | 0.635 | 0.396 | 0.577 | 0.603 |
| | Cov. Rate | 0.497 | 0.82 | 0.923 | 0.886 | 0.629 | 0.948 | 0.917 | 0.946 | 0.926 |
| | B. Cov. Rate | | | | | | 0.95 | 0.923 | 0.976 | |
| $N = 500$ | | | | | | | | | | |
| | median bias | 0.154 | 0.0669 | 0.06 | 0.0937 | 0.123 | 0.003 | 0.0266 | 0.00699 | 0.00376 |
| | Dec. Reg. | 0.266 | 0.337 | 0.37 | 0.355 | 0.338 | 0.416 | 0.358 | 0.404 | 0.416 |
| | MAD | 0.155 | 0.101 | 0.106 | 0.118 | 0.138 | 0.103 | 0.0949 | 0.102 | 0.103 |
| | RMTSE | 0.185 | 0.148 | 0.165 | 0.163 | 0.174 | 0.164 | 0.143 | 0.161 | 0.216 |
| | Cov. Rate | 0.673 | 0.881 | 0.916 | 0.883 | 0.762 | 0.959 | 0.942 | 0.958 | 0.951 |
| | B. Cov. Rate | | | | | | 0.958 | 0.942 | 0.97 | |
| $c = 0.9$ | | | | | | | | | | |
| $N = 100$ | | | | | | | | | | |
| | median bias | 0.571 | 0.234 | 0.196 | 0.148 | 0.347 | -0.0112 | 0.114 | 0.00233 | 0.00466 |
| | Dec. Reg. | 0.295 | 0.965 | 1.3 | 0.849 | 0.598 | 1.15 | 0.828 | 1.08 | 1.15 |
| | MAD | 0.571 | 0.314 | 0.323 | 0.255 | 0.356 | 0.229 | 0.231 | 0.224 | 0.238 |
| | RMTSE | 0.581 | 0.496 | 0.555 | 0.401 | 0.414 | 0.487 | 0.396 | 0.468 | 0.553 |
| | Cov. Rate | 0.0166 | 0.717 | 0.774 | 0.838 | 0.531 | 0.945 | 0.865 | 0.938 | 0.95 |
| | B. Cov. Rate | | | | | | 0.932 | 0.854 | 0.947 | |
| $N = 500$ | | | | | | | | | | |
| | median bias | 0.276 | 0.0919 | 0.103 | 0.0859 | 0.133 | 0.00263 | 0.0265 | 0.00365 | 0.033 |
| | Dec. Reg. | 0.21 | 0.339 | 0.33 | 0.418 | 0.37 | 0.367 | 0.352 | 0.364 | * |
| | MAD | 0.276 | 0.121 | 0.127 | 0.134 | 0.15 | 0.0933 | 0.0936 | 0.0935 | 0.113 |
| | RMTSE | 0.283 | 0.163 | 0.174 | 0.179 | 0.188 | 0.148 | 0.141 | 0.147 | 0.568 |
| | Cov. Rate | 0.175 | 0.842 | 0.826 | 0.884 | 0.783 | 0.959 | 0.938 | 0.958 | 0.96 |
| | B. Cov. Rate | | | | | | 0.948 | 0.928 | 0.954 | |

Table 6: Monte Carlo results

| model(c) | $R_f^2 = 0.01$ | TSLS | DNTSLS | HP | OSTSLS | STSLS | LIML | DNLIML | SLIML | REQML |
|---|---|---|---|---|---|---|---|---|---|---|
| $c = 0.1$ | | | | | | | | | | |
| $N = 100$ | | | | | | | | | | |
| | median bias | 0.0927 | 0.0721 | 0.0773 | 0.0888 | 0.0852 | 0.302 | 0.107 | 0.254 | -0.0469 |
| | Dec. Reg. | 0.568 | 3.2 | 4.68 | 0.752 | 0.863 | 20.6 | 2.13 | 11.3 | 7.9 |
| | MAD | 0.164 | 0.493 | 0.762 | 0.196 | 0.207 | 0.859 | 0.44 | 0.787 | 0.853 |
| | RMTSE | 0.243 | 0.833 | 0.957 | 0.311 | 0.402 | 0.995 | 0.735 | 0.97 | 0.992 |
| | Cov. Rate | 0.945 | 0.99 | 0.999 | 0.962 | 0.961 | 0.99 | 0.993 | 0.991 | 0.907 |
| | B. Cov. Rate | | | | | | 0.909 | 0.967 | 0.997 | |
| $N = 500$ | | | | | | | | | | |
| | median bias | 0.0882 | 0.0459 | 0.0486 | 0.0787 | 0.0795 | 0.132 | 0.0656 | 0.12 | -0.0237 |
| | Dec. Reg. | 0.482 | 1.35 | 2.79 | 0.5 | 0.631 | 4.29 | 1.25 | 3.73 | 3.06 |
| | MAD | 0.142 | 0.288 | 0.496 | 0.145 | 0.165 | 0.594 | 0.297 | 0.532 | 0.534 |
| | RMTSE | 0.206 | 0.594 | 0.81 | 0.213 | 0.305 | 0.873 | 0.532 | 0.835 | 0.833 |
| | Cov. Rate | 0.936 | 0.984 | 0.996 | 0.941 | 0.952 | 0.988 | 0.986 | 0.986 | 0.919 |
| | B. Cov. Rate | | | | | | 0.938 | 0.973 | 0.996 | |
| $c = 0.5$ | | | | | | | | | | |
| $N = 100$ | | | | | | | | | | |
| | median bias | 0.476 | 0.396 | 0.37 | 0.399 | 0.451 | 0.292 | 0.376 | 0.291 | -0.0068 |
| | Dec. Reg. | 0.507 | 2.83 | 4.31 | 2.1 | 0.777 | 5.41 | 1.89 | 4.77 | * |
| | MAD | 0.476 | 0.614 | 0.778 | 0.584 | 0.464 | 0.864 | 0.55 | 0.806 | 0.853 |
| | RMTSE | 0.514 | 0.854 | 0.957 | 0.801 | 0.561 | 0.992 | 0.776 | 0.97 | 0.989 |
| | Cov. Rate | 0.323 | 0.823 | 0.957 | 0.952 | 0.476 | 0.92 | 0.9 | 0.918 | 0.913 |
| | B. Cov. Rate | | | | | | 0.885 | 0.887 | 0.961 | |
| $N = 500$ | | | | | | | | | | |
| | median bias | 0.416 | 0.257 | 0.181 | 0.255 | 0.37 | 0.113 | 0.229 | 0.124 | -0.00747 |
| | Dec. Reg. | 0.422 | 1.34 | 2.71 | 1.13 | 0.67 | 3.27 | 1.15 | 2.84 | 6.74 |
| | MAD | 0.416 | 0.364 | 0.511 | 0.351 | 0.385 | 0.557 | 0.34 | 0.502 | 0.536 |
| | RMTSE | 0.45 | 0.623 | 0.806 | 0.515 | 0.453 | 0.846 | 0.548 | 0.81 | 0.829 |
| | Cov. Rate | 0.28 | 0.782 | 0.952 | 0.892 | 0.462 | 0.93 | 0.904 | 0.93 | 0.926 |
| | B. Cov. Rate | | | | | | 0.908 | 0.901 | 0.962 | |
| $c = 0.9$ | | | | | | | | | | |
| $N = 100$ | | | | | | | | | | |
| | median bias | 0.855 | 0.702 | 0.606 | 0.614 | 0.811 | 0.142 | 0.556 | 0.2 | 0.0511 |
| | Dec. Reg. | 0.276 | 2.36 | 3.42 | 2.21 | 0.579 | * | * | * | * |
| | MAD | 0.855 | 0.779 | 0.78 | 0.701 | 0.811 | 0.826 | 0.719 | 0.786 | 0.856 |
| | RMTSE | 0.862 | 0.91 | 0.952 | 0.884 | 0.83 | 0.98 | 0.859 | 0.956 | 1.01 |
| | Cov. Rate | 0.0004 | 0.553 | 0.739 | 0.754 | 0.125 | 0.765 | 0.604 | 0.752 | 0.94 |
| | B. Cov. Rate | | | | | | 0.542 | 0.48 | 0.788 | |
| $N = 500$ | | | | | | | | | | |
| | median bias | 0.747 | 0.395 | 0.255 | 0.284 | 0.568 | 0.0202 | 0.262 | 0.0443 | 0.0841 |
| | Dec. Reg. | 0.252 | 1.66 | 2.56 | 1.39 | 0.643 | 4.35 | 1.49 | 3.19 | * |
| | MAD | 0.747 | 0.509 | 0.5 | 0.39 | 0.569 | 0.393 | 0.377 | 0.37 | 0.531 |
| | RMTSE | 0.754 | 0.705 | 0.791 | 0.615 | 0.619 | 0.713 | 0.582 | 0.686 | 0.917 |
| | Cov. Rate | 0 | 0.646 | 0.81 | 0.788 | 0.301 | 0.895 | 0.778 | 0.895 | 0.954 |
| | B. Cov. Rate | | | | | | 0.815 | 0.725 | 0.893 | |

Table 7: Monte Carlo results

| model(d) | $R_f^2 = 0.1$ | TSLS | DNTSLS | HP | OSTSLS | STSLS | LIML | DNLIML | SLIML | REQML |
|---|---|---|---|---|---|---|---|---|---|---|
| $c = 0.1$ | | | | | | | | | | |
| $N = 100$ | | | | | | | | | | |
| | median bias | 0.0641 | 0.0734 | 0.114 | 0.0645 | 0.0659 | 0.0397 | 0.064 | 0.0419 | -0.0367 |
| | Dec. Reg. | 0.465 | 2.97 | 5.7 | 0.467 | 0.579 | 1.86 | 1.29 | 1.8 | 5.84 |
| | MAD | 0.129 | 0.349 | 0.895 | 0.129 | 0.151 | 0.348 | 0.289 | 0.345 | 0.537 |
| | RMTSE | 0.196 | 0.782 | 1.01 | 0.197 | 0.274 | 0.671 | 0.56 | 0.662 | 0.879 |
| | Cov. Rate | 0.939 | 0.98 | 1 | 0.939 | 0.944 | 0.983 | 0.985 | 0.984 | 0.921 |
| | B. Cov. Rate | | | | | | 0.959 | 0.975 | 0.996 | |
| $N = 500$ | | | | | | | | | | |
| | median bias | 0.031 | 0.0297 | 0.092 | 0.0308 | 0.0308 | 0.00119 | 0.00749 | 0.0011 | 0.000758 |
| | Dec. Reg. | 0.294 | 0.305 | 5.45 | 0.293 | 0.296 | 0.443 | 0.415 | 0.446 | 0.537 |
| | MAD | 0.0812 | 0.0835 | 0.907 | 0.0813 | 0.0811 | 0.112 | 0.108 | 0.112 | 0.122 |
| | RMTSE | 0.121 | 0.174 | 1.01 | 0.121 | 0.121 | 0.187 | 0.171 | 0.188 | 0.436 |
| | Cov. Rate | 0.941 | 0.944 | 0.998 | 0.941 | 0.94 | 0.959 | 0.956 | 0.959 | 0.945 |
| | B. Cov. Rate | | | | | | 0.958 | 0.957 | 0.979 | |
| $c = 0.5$ | | | | | | | | | | |
| $N = 100$ | | | | | | | | | | |
| | median bias | 0.317 | 0.36 | 0.457 | 0.338 | 0.324 | 0.0246 | 0.255 | 0.0342 | -0.0114 |
| | Dec. Reg. | 0.423 | 2.66 | 4.88 | 0.573 | 0.529 | 1.63 | 1.26 | 1.63 | * |
| | MAD | 0.317 | 0.499 | 0.925 | 0.34 | 0.328 | 0.323 | 0.345 | 0.321 | 0.477 |
| | RMTSE | 0.359 | 0.823 | 1.01 | 0.415 | 0.405 | 0.64 | 0.596 | 0.636 | 0.844 |
| | Cov. Rate | 0.495 | 0.746 | 0.972 | 0.63 | 0.525 | 0.942 | 0.905 | 0.942 | 0.928 |
| | B. Cov. Rate | | | | | | 0.95 | 0.915 | 0.974 | |
| $N = 500$ | | | | | | | | | | |
| | median bias | 0.154 | 0.206 | 0.416 | 0.16 | 0.158 | 0.000656 | 0.0239 | 0.00035 | 0.00626 |
| | Dec. Reg. | 0.269 | 0.953 | 4.86 | 0.283 | 0.267 | 0.423 | 0.401 | 0.425 | 0.448 |
| | MAD | 0.156 | 0.232 | 0.926 | 0.161 | 0.158 | 0.108 | 0.106 | 0.108 | 0.111 |
| | RMTSE | 0.184 | 0.574 | 1.01 | 0.194 | 0.188 | 0.178 | 0.166 | 0.179 | 0.317 |
| | Cov. Rate | 0.67 | 0.723 | 0.979 | 0.663 | 0.66 | 0.958 | 0.945 | 0.959 | 0.949 |
| | B. Cov. Rate | | | | | | 0.957 | 0.945 | 0.976 | |
| $c = 0.9$ | | | | | | | | | | |
| $N = 100$ | | | | | | | | | | |
| | median bias | 0.573 | 0.735 | 0.813 | 0.643 | 0.588 | 0.0028 | 0.371 | 0.0113 | 0.0654 |
| | Dec. Reg. | 0.291 | 2.05 | 2.92 | 0.578 | 0.348 | 1.21 | 1.14 | 1.21 | 1.72 |
| | MAD | 0.573 | 0.818 | 0.938 | 0.643 | 0.588 | 0.232 | 0.436 | 0.231 | 0.27 |
| | RMTSE | 0.584 | 0.952 | 1.01 | 0.694 | 0.612 | 0.502 | 0.653 | 0.5 | 0.629 |
| | Cov. Rate | 0.0154 | 0.558 | 0.779 | 0.221 | 0.0558 | 0.94 | 0.676 | 0.939 | 0.95 |
| | B. Cov. Rate | | | | | | 0.927 | 0.64 | 0.948 | |
| $N = 500$ | | | | | | | | | | |
| | median bias | 0.277 | 0.803 | 0.805 | 0.312 | 0.302 | -0.00076 | 0.0232 | -0.00084 | 0.0209 |
| | Dec. Reg. | 0.205 | 2.78 | 2.99 | 0.267 | 0.224 | 0.381 | 0.365 | 0.381 | * |
| | MAD | 0.277 | 0.908 | 0.935 | 0.312 | 0.302 | 0.0953 | 0.0968 | 0.0956 | 0.109 |
| | RMTSE | 0.282 | 1.01 | 1.02 | 0.335 | 0.313 | 0.158 | 0.15 | 0.158 | 0.489 |
| | Cov. Rate | 0.188 | 0.737 | 0.799 | 0.17 | 0.154 | 0.963 | 0.946 | 0.963 | 0.957 |
| | B. Cov. Rate | | | | | | 0.953 | 0.934 | 0.96 | |

Table 8: Monte Carlo results

| model(d) | $R_f^2 = 0.01$ | TSLS | DNTSLS | HP | OSTSLS | STSLS | LIML | DNLIML | SLIML | REQML |
|---|---|---|---|---|---|---|---|---|---|---|
| $c = 0.1$ | | | | | | | | | | |
| $N = 100$ | | | | | | | | | | |
| | median bias | 0.0934 | 0.105 | 0.117 | 0.0914 | 0.0971 | 0.303 | 0.154 | 0.281 | -0.0889 |
| | Dec. Reg. | 0.566 | 4.36 | 6.2 | 0.655 | 0.863 | 17 | 2.52 | 10.8 | 44.4 |
| | MAD | 0.163 | 0.612 | 0.954 | 0.18 | 0.216 | 0.852 | 0.485 | 0.803 | 0.988 |
| | RMTSE | 0.244 | 0.9 | 1.02 | 0.276 | 0.399 | 0.993 | 0.773 | 0.975 | 1.03 |
| | Cov. Rate | 0.942 | 0.992 | 0.999 | 0.951 | 0.957 | 0.989 | 0.992 | 0.99 | 0.907 |
| | B. Cov. Rate | | | | | | 0.91 | 0.957 | 0.996 | |
| $N = 500$ | | | | | | | | | | |
| | median bias | 0.0834 | 0.097 | 0.104 | 0.0803 | 0.0887 | 0.144 | 0.126 | 0.127 | -0.0441 |
| | Dec. Reg. | 0.494 | 4.04 | 6.28 | 0.497 | 0.709 | 4.71 | 1.76 | 4.11 | 7.97 |
| | MAD | 0.141 | 0.551 | 0.974 | 0.142 | 0.184 | 0.583 | 0.4 | 0.575 | 0.768 |
| | RMTSE | 0.211 | 0.879 | 1.03 | 0.214 | 0.339 | 0.872 | 0.672 | 0.863 | 0.97 |
| | Cov. Rate | 0.933 | 0.986 | 1 | 0.936 | 0.951 | 0.988 | 0.989 | 0.989 | 0.918 |
| | B. Cov. Rate | | | | | | 0.93 | 0.961 | 0.994 | |
| $c = 0.5$ | | | | | | | | | | |
| $N = 100$ | | | | | | | | | | |
| | median bias | 0.474 | 0.503 | 0.524 | 0.5 | 0.477 | 0.291 | 0.478 | 0.314 | -0.0402 |
| | Dec. Reg. | 0.505 | 3.59 | 5.04 | 1.82 | 0.752 | 5.26 | 2.09 | 4.83 | * |
| | MAD | 0.474 | 0.739 | 0.963 | 0.601 | 0.484 | 0.869 | 0.624 | 0.827 | 0.993 |
| | RMTSE | 0.514 | 0.934 | 1.03 | 0.796 | 0.586 | 0.995 | 0.832 | 0.981 | 1.04 |
| | Cov. Rate | 0.324 | 0.828 | 0.967 | 0.94 | 0.459 | 0.919 | 0.901 | 0.918 | 0.915 |
| | B. Cov. Rate | | | | | | 0.888 | 0.889 | 0.959 | |
| $N = 500$ | | | | | | | | | | |
| | median bias | 0.421 | 0.444 | 0.5 | 0.442 | 0.426 | 0.116 | 0.4 | 0.125 | -0.0266 |
| | Dec. Reg. | 0.428 | 3.54 | 5.13 | 0.938 | 0.638 | 3.41 | 1.57 | 3.29 | * |
| | MAD | 0.421 | 0.684 | 0.972 | 0.453 | 0.431 | 0.568 | 0.506 | 0.555 | 0.752 |
| | RMTSE | 0.451 | 0.916 | 1.03 | 0.579 | 0.51 | 0.848 | 0.732 | 0.844 | 0.96 |
| | Cov. Rate | 0.284 | 0.799 | 0.986 | 0.768 | 0.398 | 0.924 | 0.891 | 0.924 | 0.923 |
| | B. Cov. Rate | | | | | | 0.897 | 0.878 | 0.955 | |
| $c = 0.9$ | | | | | | | | | | |
| $N = 100$ | | | | | | | | | | |
| | median bias | 0.855 | 0.889 | 0.896 | 0.887 | 0.86 | 0.152 | 0.766 | 0.21 | 0.136 |
| | Dec. Reg. | 0.277 | 1.84 | 2.54 | 1.34 | 0.396 | * | * | * | * |
| | MAD | 0.855 | 0.93 | 0.982 | 0.894 | 0.86 | 0.84 | 0.887 | 0.825 | 1.02 |
| | RMTSE | 0.863 | 1 | 1.03 | 0.957 | 0.88 | 0.984 | 0.944 | 0.974 | 1.07 |
| | Cov. Rate | 0.0004 | 0.502 | 0.707 | 0.659 | 0.0348 | 0.763 | 0.489 | 0.745 | 0.937 |
| | B. Cov. Rate | | | | | | 0.535 | 0.37 | 0.788 | |
| $N = 500$ | | | | | | | | | | |
| | median bias | 0.748 | 0.829 | 0.897 | 0.815 | 0.757 | 0.00669 | 0.625 | 0.0226 | 0.128 |
| | Dec. Reg. | 0.261 | 1.93 | 2.66 | 0.807 | 0.361 | 5.49 | 1.51 | 5.55 | 7.23 |
| | MAD | 0.748 | 0.894 | 0.992 | 0.815 | 0.757 | 0.402 | 0.725 | 0.406 | 0.5 |
| | RMTSE | 0.757 | 0.995 | 1.04 | 0.874 | 0.782 | 0.733 | 0.811 | 0.73 | 0.862 |
| | Cov. Rate | 0.0004 | 0.513 | 0.736 | 0.338 | 0.0252 | 0.896 | 0.555 | 0.892 | 0.945 |
| | B. Cov. Rate | | | | | | 0.803 | 0.477 | 0.897 | |